Timbral constancy and compensation for spectral distortion caused by loudspeaker and room acoustics

Cleopatra Diana Pike

Thesis submitted for the degree of Doctor of Philosophy

Institute of Sound Recording,
Department of Music and Sound Recording,
University of Surrey.
2015
Originality Statement

This thesis and the work to which it refers are the results of my own efforts. Any ideas, data, images or text resulting from the work of others (whether published or unpublished) are fully identified as such within the work and attributed to their originator in the text, bibliography or in footnotes. This thesis has not been submitted in whole or in part for any other academic degree or professional qualification. I agree that the University has the right to submit my work to the plagiarism detection service TurnitinUK for originality checks. Whether or not drafts have been so-assessed, the University reserves the right to require an electronic version of the final document (as submitted) for assessment as above.

The data underlying the findings presented in this thesis are available from doi: 10.5281/zenodo.18966.

Further project information can be found at http://iosr.uk/adaptation.
Abstract

On its path to the ear the spectral envelope of a sound is modified by transmission channels. These modifications can distort, or *colour*, sound timbre preventing recognition. Research by Olive et al. (1995) suggests that perceptual mechanisms remove spectral colouration caused by loudspeakers and rooms. This compensation is apparent when listening in the real-world but not in laboratory tests and may be a result of the longer time courses involved in real-world listening. Experiments conducted as part of this thesis confirm that compensation for loudspeaker/room colouration occurs using a *real-world listening* experimental paradigm and is perceptually *moderate* to *large*. This is partly caused by mechanisms that are sensitive to the time gaps between hearing different loudspeakers/rooms, common in real-world listening, and partly due to mechanisms that are not sensitive to these time gaps. A research process was set out to further investigate mechanisms behind the time-gap sensitive component of real-world compensation.

A literature review of mechanisms that might explain this compensation was undertaken. A peripheral *enhancement* mechanism and a central *spectral compensation* mechanism cause compensation for spectral distortion caused by vocal tract (VT) characteristics in speech perception. These mechanisms are time-gap sensitive and were shown to have a number of features that mean they have the potential to cause the time-gap sensitive real-world compensation for loudspeakers and rooms. However, mechanisms that compensate for VT when listening to speech may not compensate for other channels when listening to non-speech. Laboratory tests were conducted to show that enhancement and spectral compensation also occur with non-speech sounds and therefore have the potential to contribute to any time-gap sensitive compensation for loudspeakers/rooms when listening to non-speech as well as speech. Therefore, these mechanisms can explain the real-world compensation seen in Olive et al.’s (1995) work and in real-world listening more generally. So far the specific mechanisms of real-world compensation have only been measured using laboratory studies. A framework is proposed for future work to confirm that these mechanisms explain real-world compensation using the real-world listening paradigm.
Acknowledgements

Thanks to my supervisors Dr Russell Mason and Dr Tim Brookes for all their advice and the time they devoted to this work. Thanks to Dr Chris Hope and Dr Amy Beeston for proof reading and commenting on previous versions and special thanks to Steven Steer and the rest of my family for support throughout. Thank you to all my colleagues at the Institute of Sound Recording for their help and participation in listening tests and to the many Tonmeister students who also participated in listening tests.

I am extremely grateful to the Marion Redfearn Trust and Engineering and Physical Sciences Research Council (EPSRC) who funded this work.
Contents

1 Spectral compensation and timbral constancy introduction 1
  1.1 Research questions ........................................... 3
  1.2 Chapter questions ........................................... 4
  1.3 Timbre and timbral constancy ................................ 4
    1.3.1 The definition of timbre ................................ 5
    1.3.2 Physical correlates of timbre ............................ 5
    1.3.3 Timbre in speech ........................................ 7
    1.3.4 Timbre and the hearing system ........................... 10
    1.3.5 Terminology of the hearing process ....................... 11
    1.3.6 Timbre and transmission channels ....................... 12
    1.3.7 What is timbral constancy? ............................... 16
  1.4 Thesis overview ............................................ 19
  1.5 Chapter summary and discussion .............................. 21
    1.5.1 Chapter questions ....................................... 21
  1.6 Chapter conclusion ........................................... 23

2 Compensation for loudspeaker and room timbre 25
  2.1 Assessing room and loudspeaker timbre ....................... 28
  2.2 Compensation for loudspeaker and room acoustics ............ 29
    2.2.1 Gilford (1979) ........................................... 29
    2.2.2 Olive et al. (1994) ..................................... 32
    2.2.3 Olive et al. (1995) ..................................... 34
    2.2.4 Olive and Martens (2007) ................................. 37
    2.2.5 Summary and discussion of compensation for loudspeaker and
          room acoustics ........................................... 39
2.3 Determining the best method of measuring compensation for loudspeakers and rooms

2.4 Chapter summary and discussion

2.4.1 Chapter questions

2.5 Chapter conclusion

3 Further confirmation of channel compensation in real-world listening

3.1 Experiment 1: Replicating and extending Olive et al., 1995

3.1.1 Research plan and hypotheses

3.2 Experiment 1: Overview and materials

3.2.1 Participants

3.2.2 Rating scales

3.2.3 Stimuli

3.2.4 User interface

3.2.5 Instructions

3.3 Experiment 1: Procedure

3.3.1 Loudspeaker condition

3.3.2 Room condition

3.4 Experiment 1: Results

3.4.1 Repeated measures ANOVA requirements and reliability

3.4.2 Descriptive measures

3.4.3 Statistical analysis

3.5 Experiment 1: Summary and discussion

3.6 Experiment 1: Conclusion

3.7 Experiment 2: Research aims and hypotheses

3.7.1 Rationale

3.7.2 Hypotheses

3.8 Experiment 2: Participants and material

3.9 Experiment 2: Instructions and procedure

3.10 Experiment 2: Results

3.10.1 Checking for reliability and assumptions of repeated measures ANOVA

3.10.2 Descriptive measures
3.10.3 Statistical analysis ........................................ 105
3.11 Experiment 2: Summary and discussion ....................... 113
3.12 Experiment 2: Conclusion .................................... 117
3.13 Experiment 3: Research aims and rationale .................... 118
3.14 Experiment 3: Hypotheses .................................. 119
3.15 Experiment 3: Participants and materials ...................... 120
3.16 Experiment 3: Procedure .................................... 120
3.17 Experiment 3: Results ...................................... 121
  3.17.1 Descriptive measures .................................. 121
  3.17.2 Statistical analysis .................................... 130
3.18 Experiment 3: Summary and discussion ....................... 133
3.19 Experiment 3: Conclusion .................................. 136
3.20 Chapter summary and discussion .............................. 137
  3.20.1 Chapter questions ..................................... 138
3.21 Chapter conclusion ........................................ 140

4 Testing the mechanisms of compensation for loudspeakers and rooms142
  4.1 A two-step selection and elimination process .................. 143
  4.2 Research process skeleton plan ................................ 148
  4.3 Chapter summary and discussion .............................. 151
    4.3.1 Chapter questions ..................................... 152
  4.4 Chapter conclusion ........................................ 154

5 Potential mechanisms of real-world compensation: speech and non-
  speech .................................................................. 155
  5.1 Literature review part 1: Compensation for the vocal tract and phonetic
    context ......................................................... 158
    5.1.1 Phoneme perception and vocal tract colouration ............ 160
    5.1.2 Intrinsic processes—target and modified-target theories .... 162
    5.1.3 Extrinsic processes—adapting using spectral and temporal context172
    5.1.4 Categorical perception and categorisation .................... 183
    5.1.5 Section summary and discussion ............................ 197
    5.1.6 Section conclusion ....................................... 206
5.2 Literature review part 2: General compensation for the channel in speech perception ................................. 208
  5.2.1 The enhancement effect ........................................... 210
  5.2.2 The spectral compensation effect ......................... 236
  5.2.3 Section summary and discussion ............................ 252
  5.2.4 Section conclusion .............................................. 263
5.3 Literature review part 3: General compensation for the channel in non-speech perception ..................... 264
  5.3.1 Mixed speech and non-speech tests—local context effects ... 268
  5.3.2 Mixed speech and non-speech tests—global context effects ... 275
  5.3.3 Non-speech tests .............................................. 284
  5.3.4 Section summary and discussion ............................ 290
  5.3.5 Section conclusion .............................................. 303
5.4 Establishing potential mechanisms of real-world compensation ......................................................... 304
  5.4.1 Potential compensation mechanisms: Speech .................. 306
  5.4.2 Potential compensation mechanisms: Non-speech ............. 319
  5.4.3 Non-time-gap sensitive mechanisms .......................... 324
  5.4.4 Section summary and discussion ............................ 325
5.5 Chapter Conclusion .................................................. 327

6 Experiments to test time-gap sensitive mechanisms with non-speech ..................................................... 332
  6.1 General Rationale .................................................. 335
  6.2 General hypothesis .................................................. 338
  6.3 General method .................................................... 338
    6.3.1 The brightness continuum .................................. 339
    6.3.2 Measuring shifts using the category boundary .......... 341
    6.3.3 Sensitivity and contrast ..................................... 343
  6.4 Pilot 1: Music .................................................... 346
    6.4.1 Aims .......................................................... 347
    6.4.2 Material ...................................................... 348
    6.4.3 Stimulus Production .......................................... 348
    6.4.4 Subjects ..................................................... 349
    6.4.5 Method ...................................................... 350
6.4.6 Results ................................................. 350
6.4.7 Summary and discussion .......................... 363
6.4.8 Pilot 1 Conclusion ................................. 365

6.5 Pilot 2: Noise ......................................... 367
   6.5.1 Aims .............................................. 367
   6.5.2 Material .......................................... 368
   6.5.3 Stimulus Production ............................ 368
   6.5.4 Subjects ......................................... 371
   6.5.5 Method .......................................... 371
   6.5.6 Results .......................................... 371
   6.5.7 Pilot 2 Conclusion ............................ 380

6.6 Experiment 4 ......................................... 382
   6.6.1 Aims .............................................. 382
   6.6.2 Hypotheses ...................................... 383
   6.6.3 Differences between this test and the pilots 383
   6.6.4 Material .......................................... 384
   6.6.5 The standard .................................... 384
   6.6.6 The comparison sound continuum .............. 385
   6.6.7 Precursors ...................................... 388
   6.6.8 Method .......................................... 389
   6.6.9 Subjects ......................................... 390
   6.6.10 Music-with-vocals results .................... 391
   6.6.11 Instrumental results .......................... 399
   6.6.12 Noise results .................................. 405
   6.6.13 Comparison of all 3 programme items ........ 412
   6.6.14 Experiment 4 Summary and discussion ....... 413
   6.6.15 Experiment 4 Conclusion ..................... 418

6.7 Experiment 5 ......................................... 419
   6.7.1 Participants .................................... 419
   6.7.2 Material ......................................... 420
   6.7.3 Method .......................................... 420
   6.7.4 Results—Music-with-vocals .................... 420
   6.7.5 Results—Noise .................................. 430
List of Figures

1.1 The impact of the sound modifiers (i.e the vocal tract) on the voice source to create formant peaks ........................................ 7
1.2 A spectrogram of eight vowels in the words: heed, hid, head, had, hod, hawed, hood, and who’d. Formants are indicated by arrows ............. 8
1.3 Formant frequencies for the monophthongal vowels, /i/, /e/, /u/ and /o/. 9
1.4 An illustration of the spectrum of a vowel recording in an anechoic environment and processed through two transmission channels ......... 14
1.5 The frequency spectrum of a swept sine wave played though a loudspeaker at a single within-room point and recorded at different points. .......................................................... 15

3.1 Room 1. University of Surrey, Room 09BC03b. .......................... 61
3.2 Room 2. University of Surrey, Studio Control Room 2 ............... 61
3.3 Room 3. University of Surrey, ITU BS-1116 Listening Room (TB7) . 62
3.4 Experiment 1: i) Mean preference scores for each loudspeaker in the Loudspeaker condition and ii) for each loudspeaker in the Room condition. 70
3.5 Experiment 1: i) Main effects for the loudspeaker factor in the Loudspeaker condition and ii) in the Room condition. ................. 72
3.6 Experiment 1: i) Mean preference scores for each room in the Room condition and ii) for each room in the Loudspeaker condition. .... 73
3.7 Experiment 1: i) Main effects for the room factor in the Room condition and ii) the Loudspeaker condition. ......................... 73
3.8 Experiment 2: loudspeaker effects in the the Loudspeaker condition (A) and the Room condition (B). .............................. 101
3.9 Experiment 2: main effects for the loudspeaker factor in the Loudspeaker condition (A) and the Room condition (B) ........................................ 102
3.10 Experiment 2: room effects in the Loudspeaker condition (A) and the Room condition (B). ................................................................. 103
3.11 Experiment 2: main effects for the room factor in the Loudspeaker condition (A) and in the Room condition (B). .............................. 104
3.12 Experiment 2: main effect for the loudspeaker factor (A) and room factor (B). .......................................................... 105
3.13 Experiment 3: effect of loudspeaker and room in the Loudspeaker and Room conditions for the 10minute and 0minute conditions. .... 123
3.14 Experiment 3: main effect of loudspeaker and room in the Loudspeaker and Room conditions, for the 10minute and 0minute conditions. 125
3.15 Experiment 3: effect of loudspeaker and room in the Loudspeaker and Room conditions for the Instructions and No Instructions condition. . 128
3.16 Experiment 3: main effect of loudspeaker and room for the Instructions and No instructions conditions ........................................ 129

5.1 Spectrogram patterns sufficient for the synthesis of /d/ before vowels. . 161
5.2 F1 and F2 formant frequency values of American English vowels. Koenig frequency scale used. ................................................................. 163
5.3 Taken from Johnson (2005). Spectrogram of a man and a woman saying ‘cat’. .......................................................... 164
5.4 F1 and F2 locations for American English Vowels, for a variety of speakers.166
5.5 Categorical perception. F2 locus of origin below 1500 Hz results in /b/, 1500-2000 Hz, /d/ and above 2000 Hz, /g/ identifications. ............... 185
5.6 Categorical perception and memory—consonants and vowels: identification and discrimination performance and discrimination as a function of stimulus pair and delay. ........................................ 190
5.7 Categorical perception and memory—sawtooth stimuli: identification and discrimination performance and discrimination performance as a function of stimulus pair and delay. ......................... 191
5.8 Sine-wave speech, ‘where were you a year ago’: narrowband spectrogram (a), wideband spectrogram (b) and narrowband spectrogram of 3 sinusoidal replicas (c). ........................................ 194

5.9 The frequency spectrum of a harmonic complex cued by that same complex and the corresponding perception. ....................... 211

5.10 FSV tests: Flat spectrum precursors have harmonics representing formant frequencies reduced. Harmonics are reintroduced in test sounds. Vowels with spectra that are the inverse of the precursor are heard. ... 213

5.11 Decremental FSVs and Decremental PSVs show enhancement of a vowel when components not making up the vowel are decreased in amplitude between the precursor and test stimulus ........................................ 214

5.12 PSV tests. Precursor: a uniform spectrum consisting of 50 harmonics at equal amplitude. Test stimulus: a vowel created by amplifying pairs of harmonics close to the first 3 formants of an ‘AH’ vowel. ....................... 215

5.13 The distorting effect of transmission channels: a single vowel sound in anechoic conditions ‘in quiet’ and the same vowel produced through two different channels with different frequency responses. ................. 216

5.14 Description of FSV, PSV and DSV enhancement effect tests .......... 217

5.15 Effect of amplitude contrast on the enhancement effect ............. 219

5.16 Gain in vowel identification accuracy between enhanced and non-enhanced PSV vowels with dB contrast and confidence of judgements. 220

5.17 Magnitude of enhancement effect in an FSV experiment as a function of the duration of the vowel-complement precursor. ................. 222

5.18 Recovery of enhancement measured by a decrease in identification accuracy as a function of transition length and silent time gap. ...... 223

5.19 Enhancement and ‘simple’ neural adaptation .......................... 230

5.20 Watkins (1991): the enhancement effect and the spectral compensation effect as measured by shifts in the category boundary of an ‘Itch’ ot ‘Etch’ continuum with different precursors ............................ 237

5.21 Watkins (1991): the enhancement effect and the spectral compensation effect. Precursor and test played to same ear, to different ears or to same ears but heard from different directions. ......................... 239
6.1 Pilot 1: sensitivity with the music-with-vocals track .......................... 352
6.2 Music-with-vocals track—sensitivity with monaural listening ............... 359
6.3 Pilot 2: sensitivity with the noise programme item ............................. 373
6.4 Noise track—sensitivity with monaural listening ................................ 376
6.5 The proportion of ‘More’ HF than standard responses for each S/C pair for the music-with-vocals programme item ................................. 391
6.6 The difference in the proportion of ‘More’ HF than standard responses with ‘More HF’ and ‘Less HF’ precursors—music-with-vocals stimuli .... 398
6.7 The proportion of ‘More HF than standard’ responses for each S/C pair for the instrumental programme item ........................................... 400
6.8 The difference in the proportion of ‘More’ HF than standard responses between ‘More HF’ and ‘Less HF’ precursors—instrumental stimuli .... 405
6.9 The proportion of ‘More’ HF than standard responses for each S/C pair for the noise programme item .................................................... 406
6.10 The difference in the proportion of ‘More’ HF than standard responses between ‘More’ and ‘Less’ HF precursors—noise stimuli ................. 412
6.11 Difference in proportion of ‘More HF’ responses between the ‘More’ and ‘Less’ HF precursors, across all S/C intervals, for all programme items . 413
6.12 The proportion of ‘More’ HF than standard responses at each S/C interval for the music-with-vocals crossaural stimuli .............................. 421
6.14 The proportion of ‘More’ HF than standard responses at each S/C interval for the noise stimuli with crossaural precursor presentation ....... 431
6.15 Difference in the proportion of ‘More HF’ responses at each S/C interval between ‘More’ and ‘Less’ HF precursors, for the noise stimuli: crossaural presentation ........................................ 437

B.1 Recording apparatus ................................................................. 520

C.1 Task interface ................................................................. 521

D.1 Photograph: Room 1 ......................................................... 522
D.2 Photograph: Room 2. ............................................ 522
D.3 Photograph: Room 3. ............................................ 523

E.1 Reliability analysis of Participant 1. Cronbach’s alpha = .789 .... 525
E.2 Reliability analysis of Participant 2. Cronbach’s alpha = .264 .... 525
E.3 Reliability analysis of Participant 3. Cronbach’s alpha = .246 .... 526
E.4 Reliability analysis of Participant 4. Cronbach’s alpha = .823 .... 526
E.5 Reliability analysis of Participant 5. Cronbach’s alpha = .001 .... 527
E.6 Reliability analysis of Participant 6. Cronbach’s alpha = .771 .... 527
E.7 Reliability analysis of Participant 7. Cronbach’s alpha = .864 .... 528
E.8 Reliability analysis of Participant 8. Cronbach’s alpha = .835 .... 528
List of Tables

1.1 The chain of stimulus to perception. Red text shows central processes and blue text shows peripheral processes. 11

3.1 Experiment 1 hypotheses 52
3.2 Objective characteristics of loudspeakers used in Experiment 1. 59
3.3 Typical order of presentation of stimuli for a participant completing the Loudspeaker condition and Room condition 24 h later 66
3.4 Loudspeaker mean ratings within each room and loudspeaker mean variance within each room. 71
3.5 Room means within each loudspeaker and room variance within each loudspeaker. 74
3.6 Anova analysis: Main effects and loudspeaker*room interactions for each factor in the loudspeaker and room conditions. 76
3.7 Variance measures used to calculate main effect of loudspeakers and rooms when compared directly. 78
3.8 Main effect by condition interactions 79
3.9 Familiarisation Phase. Stimulus presentation order is shown for a single participant. 96
3.10 The procedure (both conditions) for a single participant starting in the loudspeaker condition. 97
3.11 Time spent listening to the indirect factor. 99
3.12 Loudspeaker means within each room, loudspeaker variance within each room. 102
3.13 Room means for each loudspeaker and room variance for each loudspeaker. 104
3.14 Separate repeated measures of ANOVA analysis by condition. Data from experimental pages 1-9. .............................. 107
3.15 Separate repeated measures of ANOVA analysis by condition. Data from experimental pages 1-3. .............................. 108
3.16 Separate repeated measures of ANOVA analysis by condition. Data from experimental pages 7-9. .............................. 109
3.17 Single repeated measures of ANOVA analysis, both conditions. ..... 110
3.18 Loudspeaker means within each room and loudspeaker mean variance within each room, in the 0 minute time gap condition. ............... 122
3.19 Room means for each loudspeaker and room mean variance for each loudspeaker. In the 0 minute time gap condition ............... 122

5.1 An example of the frequencies omitted or reduced from a flat spectrum to create a vowel complement precursor stimulus. ................ 212

6.1 Category boundary measures of compensation. ...................... 344
6.2 Gain values (dB) for filtering above 1 kHz: absolute difference between standard and comparison sound and exact dB gain filtering values. . 349
6.3 Probability of a ‘More’ HF than the standard response for the music-with-vocals continuum. .............................................. 353
6.4 Probability of a ‘More’ response for the music-with-vocals continuum with monaural comparisons. ...................................... 360
6.5 Pilot 1: Bias measures for the music-with-vocals continuum .......... 362
6.6 Gain values (dB) for filtering above 1 kHz: absolute difference between standard and comparison sound and exact dB gain filtering values. . 370
6.8 Probability of a ‘More’ HF response for the Noise continuum with monaural comparisons. ...................................... 377
6.9 Bias measures for the Noise programme continuum ............... 378
6.10 Gain values (dB) for filtering above 1 kHz. S/C pairs are formed between each comparison and the standard in the row below for the musical programme items. ................ 387
6.11 Bias measures for the HF continuum composed of the Music-with-vocals programme item ........................................ 396
6.12 Bias measures for the HF continuum composed of the Instrumental programme item ........................................ 403
6.13 Bias measures for the HF continuum composed of the Noise programme item .................................................. 410
6.14 Bias measures for the HF continuum composed of the music-with-vocals programme item with monaural comparisons and a precursor presented contralaterally .................................................. 425
6.15 Bias measures for the HF continuum composed of the Noise with the LTAS of music-with-vocals programme item with monaural comparisons and a precursor presented contralaterally ....................... 435
Chapter 1

Spectral compensation and timbral constancy introduction

The environment in which a sound is heard, or the transmission channel (e.g. microphone, loudspeaker, listening room) through which a sound is heard can affect its spectrum. A sound’s spectrum, or its perceptual correlate, timbre, is an important cue for identifying the sound. Perceptual mechanisms that maintain constancy of timbre may be necessary to ensure that a sound is recognised despite spectral modification caused by different channels. One way in which timbral constancy may be achieved is by a hearing system which compensates for the effects of the channel.

There has been a large amount of research into perceptual compensation for the effects of room reflections on the perceived location of a sound source (e.g. the Precedence Effect (Litovsky et al. 1999)) and on the intelligibility of speech (Beeston et al. 2014, Watkins 2005a, Zahorik et al. 2009). However, to date there has been relatively little research into compensation for the effects of the room or other channel on a sound’s timbre. Ensuring accurate timbral perception is likely to be important to perception but a general mechanism of compensation for spectral colouration caused by the channel has not been confirmed and the mechanisms behind such compensation are not well understood.

If it exists, this compensation process is likely to be of interest to hearing scientists as it may show both a long-evolved mechanism which ensures the perceptual stability of
auditory objects of biological significance across different environments (e.g. the sound of a predator or prey), and mechanisms that have developed during the evolution of complex language to help the listener to understand the sensitive speech signal across potentially distorting environments. It is likely that any spectral compensation mechanism is complex and occurs at multiple levels within the hearing system. A mechanism that compensates for the channel may also play a role increasing the dynamic range of the hearing system. This role in enhancing the sensitivity of the hearing system may be as important as creating stability in perception. Research into compensation for the spectral effects of the channel is also likely to be of interest to audio engineers. Room adaptive loudspeakers or headphones that remove distracting or distorting aspects of the listening environment may be more or less useful depending on the extent to which the listener already compensates for these things. Further, understanding how the human hearing system deals with spectral colouration is likely to be of interest to those developing machine listening devices.

Some research into compensation for the effects of loudspeakers and listening rooms has been conducted by researchers in the audio engineering community. However, compensation for these channels has not been confirmed and most research into the perception of loudspeakers and listening rooms is conducted in laboratory listening scenarios where compensation is not expected to occur to the extent that might occur in the real-world. As will be explained further in Chapter 2 the lack of compensation in the laboratory appears to be largely due to the shorter listening periods in laboratory listening compared to real-world listening. To better show the extent of real-world compensation, real-world experiments will be conducted as part of this thesis. These experiments are described further in Chapter 2 and primarily aim to replicate the longer listening time course that occurs in real-world listening more closely. Outside of the audio engineering domain, research into compensation for the spectral effects of the channel has been carried out almost entirely within the field of speech perception (vocal tract (VT) normalisation and compensation for the phonetic context). The extent to which the mechanisms behind compensation for the VT and phonetic context are also involved in compensation for other channels, when listening to sounds other than speech, is not clear. Therefore, if compensation for loudspeakers and rooms is shown in real-world tests, this thesis further aims to determine whether the same mechanisms
involved in compensation for the VT and phonetic context in speech perception are behind this.

The remainder of the current chapter describes the specific research questions that this thesis aims to answer (see Section 1.1). Three chapter questions for the current chapter are then set out (see Section 1.2). Section 1.3 consists of a number of subsections which aim to further describe and define the research problem and some of the terminology used in this thesis. An overview of the rest of the thesis is given in Section 1.4 and answers to the chapter questions are given in the summary and discussion section of this chapter (see Section 1.5). A conclusion to the chapter is presented in Section 1.6. The extent to which the main research questions can be answered using the information presented in this chapter is discussed here in brief and in Chapter 7.

1.1 Research questions

The research in this thesis will aim to answer the primary research question:

‘How could perceptual compensation for spectral distortion caused by the transmission channel affect the perception of loudspeakers and rooms in real-world listening?’

This can be divided into 2 research questions which will be addressed in this thesis:

**Research Question 1a:** ‘In what way can compensation for spectral distortion caused by the transmission channel affect the perception of loudspeakers and listening rooms?’

This question aims to determine the nature and extent of the effect that compensation may have on the perception of sounds heard through loudspeakers in rooms and work that aims to answer this question will be presented in Chapters 1, 2 and 3 of this thesis.

**Research Question 1b:** ‘By what means can compensation affect the perception of loudspeaker and rooms?’ This question aims to determine some of the potential
mechanisms that may be involved in compensation when listening to music and speech through channels in the real world. This question will mainly be answered in Chapters 3 to 7 of this thesis.

Another question is also of interest:

**Research Question 1c:** ‘What are the actual mechanisms of real-world compensation for loudspeaker and rooms?’

This question will not be answered directly in this thesis but the work presented leads to a research process to determine the extent to which each of the potential mechanisms actually causes compensation when listening to loudspeakers and rooms in the real-world. This mechanism will be discussed in more detail in Chapters 4 and 7.

### 1.2 Chapter questions

The following chapter questions are presented. These questions aim to guide the work within the remainder of the current chapter towards setting the context for this thesis and providing answers to Research Questions 1a and 1b.

**Question 1.1:** How is timbre used in sound recognition in speech and non-speech perception?

**Question 1.2:** What is the effect of the transmission channel on timbre and sound recognition?

**Question 1.3:** What is timbral constancy and how does this occur to reduce the effect of the channel to allow for more accurate auditory object identification?
1.3 Timbre and timbral constancy

This section further sets the context for the work in this thesis and further defines the scope of the research questions and some of the terms used within this thesis. Information is presented that will provide answers to the chapter questions and research Questions 1a and 1b.

1.3.1 The definition of timbre

Timbre can be described as a sensation of tone quality, colour or character that is produced by a musical instrument or other sound producing object. Timbre is responsible for our ability to discriminate between different sounds and identify sounds as coming from different objects (Fletcher 1934, McAdams 1993). For example a violin is distinguishable from a clarinet because of the character of the sound it produces. Timbre is not the only aspect of sound that contributes to auditory object identification; the magnitude of a sound and its directional and temporal properties may also be relevant (Bregman 1990, Kubovy and Van Valkenburg 2001) but these are less important to object identification than timbre. Formal definitions of timbre highlight the fact that timbre encompasses the residual aspects of tone sensation that are not to do with pitch or loudness (Plomp 1976). A formal definition is given by the American National Standards Institute (ANSI 1960): Timbre is ‘the attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar.’

1.3.2 Physical correlates of timbre

Timbre perception is multi-dimensional and thought to be caused by a number of physical attributes. ANSI (1960) describes the main physical correlates of timbre: ‘Timbre depends primarily upon the spectrum of the stimulus but also upon the waveform, the sound pressure, the frequency location of the spectrum and the temporal characteristics of the stimulus’. Broadly speaking, timbre is determined by time-invariant patterns (spectral content) and time-variant patterns (temporal
amplitude envelopes) (Moore 2003). Of these aspects the most commonly discussed, and possibly the most important, is the spectral composition of a sound, during the steady-state portion of the sound. The spectral character of the steady-state portion of sound largely depends on the relative amplitudes of the frequency components after modification by the instrument’s body. This is known as the spectral envelope of the sound. Musical instruments produce sounds with unique steady state spectral characteristics due to unique resonant properties of the instrument body (Toole and Olive 1988). The size, shape and material of the instrument causes different frequencies in the spectrum to be amplified and attenuated to different degrees in different instruments (e.g. the same note played on a violin or viola will sound different because of instrument body size, shape and material). Tonal instruments amplify frequencies that are integer multiples of the fundamental frequency, known as harmonics. Different instruments amplify and attenuate each harmonic differently. For example the clarinet body favours odd numbered harmonics and this pattern of amplification is responsible for much of the distinctive sound of the clarinet. The distinctive timbre of non-pitched instruments is due to the presence of components at unevenly spaced multiples of the fundamental, or non-harmonic partials. The way that the source sound is produced by the instrument can also influence its spectral properties. For example, a struck string will excite more non-harmonic partials resulting in a more percussive sound than a bowed string. Noise or non-harmonic components may also occur during the course of tonal sounds. For example, the noise that creates a perception of breathiness at the onset of a flute sound is essential for differentiating a flute sound from pure tone sounds (Howard and Angus 2009). Instrument sounds do not only consist of a steady-state component. Temporal aspects such as the attack and decay of the sounds are also important for defining timbre; ‘The recognition of musical instruments, for example, depends quite strongly on onset transients and the temporal structure of the sound envelope’ (Risset and Wessel 1999). Sounds with similar harmonics may sound different depending on the rate at which those harmonics come into effect and the amplitude of the harmonics during the onset and offset of the sound. Schouten (1968) lists several other factors that are likely to contribute to timbre and auditory object recognition:

1. whether the sound is periodic, having a tonal quality for repetition rates between
20 and 20,000 Hz or irregular, having a noise like character;

(2) whether the waveform envelope is constant or fluctuates as a function of time, and in the latter case what the fluctuations are like;

(3) whether any other aspect of the sound (e.g. spectrum or periodicity) is changing as a function of time; and

(4) what the preceding and following sounds are like.

### 1.3.3 Timbre in speech

Timbre is essential to the perception of speech. Vocal sounds are produced in the larynx by the vibration of the vocal folds. Air is passed from the lungs through a gap between the vocal folds called the glottis. This is known as the glottal source. In speech the glottal source usually has a low fundamental frequency and consists of a wideband spectrum with more energy at low frequencies (Figure 1.1, a).

![Figure 1.1](image)

*Figure 1.1:* The impact of the sound modifiers on the voice source to create formant peaks when producing an ‘ah’ vowel. (a) the unmodified glottal source, (b) modifications due to vocal tract shape applied to the source, (c) the resultant output. Taken from Howard and Murphy (2008).

The fundamental frequency of the sound can be altered by altering the rate of vocal chord vibration to produce sounds of different pitch. The sound then passes through the VT, which includes the rest of the system that lies above the larynx. The spectral shape of the sound is modified here by the tongue, lips and jaw. The VT modifies the sound source in a similar way to a musical instrument body: certain frequencies are amplified relative to others. It changes the spectral envelope of the sound and introduces resonance peaks known as formants (Figure 1.1, b). The shape of the VT
varies to produce different speech sounds or *phonemes*. The shape determines the spectral envelope of each phoneme and spectral location of its formants (Figure 1.1, c) (Moore 2003).

Source filter theory describes this process (Fant 1960). A frequency domain representation based on this is shown in Equation 1.1. The excitation source (glottal source), $e(f)$, excites the air within the vocal tract. The vocal tract is represented by a filter having a transfer function, $i(f)$. This gives the resultant sound, $s(f)$.

$$s(f) = i(f)e(f)$$

(1.1)

The properties of vocal tracts, such as size and shape also vary between speakers giving speakers differing sets of resonances contributing, along with articulatory habits, to their own distinctive speech sound. The timbre of speech is therefore also used for speaker identification as well as for recognising different phonemes. Spectrograms reveal information about the frequency content of phonemes, their temporal amplitude properties and formants (Figure 1.2). The formants for a number of vowels are shown by dark areas in the spectrogram, which show peaks in the resonance of the vocal tract at different frequency locations. The dark areas in the spectrogram of the [a] vowel in Figure 1.2 correspond to the amplitude peaks in the same vowel shown in Figure 1.1, c.

Speech research has attempted to establish all the objective acoustic properties of speech that relate to the perception of different phonemes. Formants are thought to be particularly important to speech perception. Early research focused on the nature of the simplest speech sounds, the monophthongal vowels, depicted in Figure 1.2 (these are considered to be simple because when produced in isolation they consist of a steady state sound—although it is now known that even in isolation they may be subject to fluctuation (Hillenbrand et al. 1995). For other phonemes the vocal tract shape is changed during production (e.g. diphthong vowels, plosives.) Formants were shown to be the basis for vowel identification. Helmholtz (1863) used formant frequencies to synthesize vowels and found that monophthongal vowels could be created by raising the amplitude of formants in a wideband spectrum. The frequency location of formants
1.3 TIMBRE AND TIMBRAL CONSTANCY

was found to determine the identification of different vowels in a number of studies (see Miller (1989), and Nearey (1989) for a review). Figure 1.3 demonstrates the importance of formant location. Formants 1 and 2 are widely separated in an [I] vowel, but are closer together for [O].

Figure 1.3: Formant frequencies for the monophthongal vowels, /i/, /e/, /u/ and /o/. From Ohl and Scheich (1997)

Other phonemes are also identified by their spectral content and formants peaks, but in the case of other sounds other factors play an increasingly important role including: the movement of formant frequencies during the sound (formant transitions), the timing at which certain formants begin (Voice Onset Time, VOT) and the spectral shape for
voiceless sounds where the vocal chords do not vibrate, such as in the production of [s].

That speech sounds can be recognised mainly by the frequency location of the first few formants, as well as temporal information, is the basic view of speech perception that prevails in introductory speech texts. However, there is now a large body of research showing that the situation is more complex. A number of other factors have been identified as important to the identification of phonemes including: the relative (rather than absolute) position of formants (Miller 1989, Nearey 1989); the whole spectrum of the sound or spectral tilt (Ito et al. 2001, Kiefte and Kluender 2005, Liu and Eddins 2008), the fundamental frequency, the frequency and position of the third formant and the presence of higher formants above the third formant (Miller 1989, Nearey 1989).

There is also blurring of the objective speech cues through co-articulation, meaning that the acoustical signal for each phoneme is different depending on the surrounding context (adjacent speech sounds made by the same speaker) (Liberman et al. 1967). These issues are discussed further in Chapter 5. However, despite the complex nature of the acoustical cues for most speech sounds there remains much evidence to support the view that carefully spoken monophthongal vowels are recognised by their spectrum and in particular their steady state formant frequencies. This has allowed for the use of simplified synthesised vowel stimuli in experiments such as Summerfield et al. (1984) which will be discussed further in Chapter 5.

1.3.4 Timbre and the hearing system

This section introduces the processes involved in timbre perception. A description of the processes is useful to give an insight into the possible points along the stimulus-to-perception chain that mechanisms which compensate for distortion caused by the transmission channel occur. This section also explains some of the terminology used in the research in this area.

The stimulus-to-perception chain

The stimulus-to-perception chain describes the hearing process (see Table 1.1). The ear provides the means for mechanical transduction of sound from a physical medium.
Table 1.1: The chain of stimulus to perception. Red text shows central processes and blue text shows peripheral processes.

| Stimulus | ↓ | Modification by outer ear | ↓ | Transduction by inner ear | ↓ | Peripheral neural processing | ↓ | Neural processing in higher centres | ↓ | Cognitive processing | ↓ | Semantic processing |

(changes in air pressure) to biological information (neural firing) that is then passed on by the auditory nerve to the brain. At the auditory periphery sounds at each ear are processed separately. At the cochlear nuclei sounds at both ears combine and begin to travel higher neural pathways. At these central sites the sound or parts of the sound may be subject to further physiological processes, such as adaptation of neural firing rates. The sound will then undergo interpretation by the cognitive system to result in the perception of sound timbre and then auditory object recognition. The next stage is the semantic processing stage, which involves giving the sound linguistic meaning (do the phonemes make up a word? What does the word mean?) and attaching emotional significance to sounds (e.g. fear of loud bangs, emotion in music). Processes that compensate for the channel may occur at any point in the hearing process between the transduction and the semantic stages.

1.3.5 Terminology of the hearing process

This subsection explains some of the terms used to describe hearing processes. These terms will be used in discussions in the following chapters. The key processes that are involved in timbre perception may be broadly characterized as physiological or
cognitive and peripheral or central.

**Cognitive processes** occur higher up in the stimulus-to-perception chain. The term cognitive is usually used to describe executive and intellectual functions such as memory, learning, and attention (Kellogg 2007). A key feature of many cognitive processes is that their basis is ultimately neural but their complexity means that the exact physiological mechanisms behind them are unknown. Cognitive processes are usually described as complex distinct processes that interact with each other within the *cognitive system* to achieve a goal (e.g. memory and attention must interact in order to remember an event). An analogy is often made with the functions of various *modules* within a computer system working towards a task and there may exist distinctly separate cognitive systems for different tasks (Coltheart 1999). Another key feature is the scope for conscious control. A cognitive process may be modifiable by learning and training, whereas physiological process are often automatic and difficult to modify. **Physiological processes** are processes for which the biological (mechanical, electrical, chemical, or *neural*) underpinnings are well described. These tend to be simple processes which achieve a specific simple goal but at higher centres may interact to cause a more complex cognitive process (Brown and Wallace 2012). These occur at all levels of the stimulus-to-perception chain and tend to be automatic and not subject to conscious control (unless underpinning a controllable cognitive process). At lower levels, physiological processes are particularly well understood and particularly automatic as they largely involve receiving sensory signals and passing these to higher centres for further evaluation. However, top-down *efferent* signals can affect these lower-level processes.

The terms *Peripheral* and *Central* broadly describe where in the stimulus-to-perception chain a process occurs. In speech psychophysics, the term peripheral is usually used to describe processes occurring below the cochlear nuclei. A peripheral process acts on one ear only and the function of one ear cannot affect the other directly. However, both ears may undergo the same peripheral process at the same time. The term Central is usually used in research to describe processes occurring beyond this, where stimuli from both ears meet. Central processes can involve information from both ears and can be affected by events occurring at either ear. The term central often refers to high level cognitive processes but may be used to refer to any physiological
1.3 TIMBRE AND TIMBRAL CONSTANCY

processes occurring at or later than the cochlear nuclei.

1.3.6 Timbre and transmission channels

Any object or space through which a sound passes, such as an open space, room, or an electro-acoustic channel (e.g. a microphone, loudspeaker or telephone line) can be referred to as a transmission channel (Watkins 1991). Vocal tracts, musical instrument bodies and human ears can also be regarded as transmission channels through which sound passes. All these channels modify the sound in the similar way as was described for the vocal tract in Section 1.3.3. The previously described source-filter model (Equation 1.1) can be used to describe the effects of any transmission channel on a sound’s spectrum. For example, to describe the effect of a further transmission channel, beyond the vocal tract, on speech (e.g microphone, loudspeaker or room), a further filter parameter, $j(f)$, can be added to the source filter model to account for modification by that transmission channel, where $i(f)$ is the vocal tract, $e(f)$ the excitation source and $s(f)$ the resulting sound:

$$s(f) = j(f)i(f)e(f). \quad (1.2)$$

Modification by transmission channels beyond the instrument body is frequently referred to as colouration. Colouration is often desirable and can be manipulated to enhance the timbre of instrument sounds in live and recorded listening. This is done by the choice and placement of microphones, loudspeakers and absorbing materials, and via the mixing process. Colouration can also cause unpleasant timbral effects and may be considered spectral distortion. For this reason colouration that is not controlled in the recording process or live set-up is usually unwanted. However, in this research the terms distortion and colouration are used in a value neutral way to simply describe a modification to the timbre of a sound source.

Colouration occurs due to the uneven frequency response of the rooms and loudspeakers: ‘When sounds are transmitted from a source to a listener the spectral envelope is invariably and diversely distorted by factors such as room reverberation’
1.3 TIMBRE AND TIMBRAL CONSTANCY

Coloration produced by loudspeakers is often said to detract from the fidelity of the original sound source. For this reason many loudspeakers attempt to reproduce sounds without colouration. Toole (2008) observes: ‘the task of a sound reproduction system is to accurately portray the panorama of resonances and other sounds in the original sources not to editorialise by adding its own’. Because of its apparent reliance on spectral attributes for identification, the colouration of speech by transmission channels has the potential to affect listeners’ ability to perceive speech accurately:

‘Vowels are heard in different spectral contexts. The spectral context being the frequency response of the room in which they are heard, or the frequency response of different vocal apparatus [...] if detail in the spectral pattern is degraded either externally in the communication channel or internally in the peripheral auditory system, then frequency resolution will be reduced and the identification and discrimination of speech will suffer’ (Summerfield and Assmann 1989).

The impact of the transmission channel on speech can be seen in Figure 1.4 taken from an experiment by Summerfield and Assmann (1989). This figure demonstrates the different spectral distortion when a single vowel sound is passed through two different transmission channels. The resulting spectra are spectrally dissimilar from each other and from the vowel produced anechoically. A comparable effect on the frequency spectrum occurs when any sound is produced through any transmission channel with an uneven frequency response (Summerfield and Assmann 1989).

Benade (1984) also demonstrates the problem by recording a swept sine wave (a sinusoidal signal of constant amplitude and slowly rising frequency) with the same microphone placed in different positions within a room (Figure 1.5). Benade describes ‘turmoil in the signal path’:

‘There is great irregularity in the signal that is recorded by the microphone and the pattern of irregularity is completely altered between the 2 microphone positions [...] and has no direct similarity to the source spectrum [...] if we accept the traditional dogma that the tone colours
1.3 TIMBRE AND TIMBRAL CONSTANCY

Figure 1.4: An illustration of the spectrum of a vowel recording in an anechoic environment ('in quiet') and processed through two transmission channels with different, uneven, frequency responses. Taken from Summerfield and Assmann (1989)

of sound are primarily determined by their spectra, we have apparently proved that recognizing the sound of anything in a room is an essentially hopeless task.'

In the example discussed by Benade variations are particularly large because of the presence of room modes, but transmission channels such as loudspeakers and telephones can also produce large distortions (Warren et al. 1995).

1.3.7 What is timbral constancy?

If spectral composition is key to recognising a sound source, disruption caused by the channel should make it difficult for listeners to identify sounds by their spectrum. However, humans are able to recognize sounds with ease despite transmission channel variation. Benade (1984) observes that we seldom need to give conscious thought to the problems of hearing in a room:

‘the physicist says that the signal path in a music room is the cause of great confusion whereas the musician and his audience find that without the room only music of the most elementary sort is possible! Clearly we have a paradox to resolve as we look for the features of the musical sound that
Figure 1.5: The frequency spectrum of a swept sine wave played though a loudspeaker positioned at a single point within the room and recorded at two different points within the room. Taken from Benade (1984)

gives it sufficient robustness to survive its strenuous voyage to its listeners and as we seek the features of the transmission process itself that permit a cleverly designed auditory system to deduce the nature of the sources that produced the original sound'.

Watkins (1991) states that to account for relatively constant perceptions across transmission channels a perceptual mechanism may be used to create perceptual constancy:
‘the prominent distortion of the spectral envelope occurs in both small and large rooms [...] and other communication channels in common use such as telephones [...] it is nevertheless possible to identify a sound when there is spectral envelope distortion [...]. These observations imply that there may be a mechanism for ‘timbral constancy’ (Risset and Wessel 1999) that compensates for the distortion of the channel in the perceptual assessment of the spectral envelope of the sound source’.

Although the search for a mechanism that compensates for channel spectrum has been carried out in speech research (VT and phonetic context colouration—see further Chapter 5), only a few researchers have examined timbral constancy as a general auditory process (Darwin et al. 1989, Kluender et al. 2003, Summerfield et al. 1984, 1987, Watkins 1991). The definition of timbral constancy has varied. At its most broad the investigation of timbral constancy has involved the examination of timbral stability of a sound, not only across changes in its spectral envelope due to transmission channel, but also across changes in pitch or amplitude of the source, and when heard alongside other sounds with different spectra, which temporally overlap with the sound and so may effect the spectrum of a the sound reaching the ear. Timbral constancy in this broad sense could be defined as:

‘the stability of timbre of a complex sound relative to changes in the sound’s spectral envelope’.

Risset and Wessel (1999) have used the term in its broad sense to explain the fact that the sound originating from a musical instrument is heard as timbrally similar when that instrument is playing different notes in the musical scale or when different dynamics are used:

‘a form of timbral constancy is implied by the common observation that a sound source can be reliably identified over a wide variety of circumstances, for example a saxophone is readily identified as such regardless of the pitch or dynamic, and whether it is heard over a distortion ridden pocket size transistor radio or directly in a concert hall.’ (Risset and Wessel 1999)
Bregman (1990) has used the term to describe the constancy of a sound when its spectrum is intermixed with other sound sources that are adjacent in space:

‘we can show that complex phenomena as constancies [in vision] exist in hearing too. One example is timbre constancy. A friend’s voice has the same perceived timbre in a quiet room as at a cocktail party. Yet at the party the set of frequency components arising from the voice is mixed at the listeners ear with frequency components from other sources. The total spectrum of energy that reaches the ear may be quite different in different environments. To recognise the unique timbre of the voice we have to isolate the frequency components that are responsible for it from others that are present at the same time [...] the fact that we can usually recognize timbre implies that we regularly choose the right components in different contexts. Just as in the case of the visual constancy, timbre constancy will have to be explained in terms of complicated analysis by the brain.’

However, most research that has examined timbral constancy has considered the problem to be narrower (see Chapter 5 for a description of these studies). These studies have used the term to describe the stability of timbre relative to spectral changes caused by different transmission channels. This situation is that which was described by Risset and Wessel (above), where a saxophone is readily identified whether it is heard over a distortion ridden transistor radio or directly in a concert hall. Any mention of timbral constancy in this thesis refers to timbral constancy in this narrower sense. Therefore, the definition of timbral constancy for the purpose of this thesis may be:

‘The stability of timbre, relative to changes in a sound’s spectral envelope that occur when the sound passes through different channels with different frequency responses’.

This definition does not include timbral constancy over changes in pitch, dynamics or in the face of other separate sounds sources heard simultaneously (sounds adjacent in space). However, it may be desirable to also include constancy across sounds that are temporally adjacent within this definition. This is because in speech perception sounds made by the same speaker prior to the speech sound in question (the *phonetic context*) cause spectral distortion of that sound, which is physically similar to colouration caused
by the VT (see further Chapter 5). The same mechanisms appear to be behind compensation for the channel and compensation for the effect of the phonetic context in speech perception (namely both may use a sample of prior sound to assist the compensation). Therefore, constancy in light of the prior phonetic context will be discussed.

The research definition of timbral constancy used in this report is therefore:

‘The stability of timbre, relative to changes in a sound’s spectral envelope that occur when the sound passes through different channels with different frequency responses, or is heard in different prior phonetic contexts’.

Limiting the scope of this research to timbral constancy across transmission channels and across different phonetic contexts may be warranted because it appears likely that transmission channel/phonetic context constancy involves the same mechanisms but that constancy across pitch and loudness, and constancy in light of different contemporaneous sound sources, represent different problems and involve different mechanisms. For example, a separation of sounds adjacent in space involves source segregation based on spatial grouping cues that are not available when a sound’s spectrum must be separated from that of a transmission channel such as a microphone or loudspeaker (Bregman 1990). However, it is noted that when one considers that prior listening to the channel/context may be necessary to separate the channel from the source (as will be seen in Chapter 5), grouping cues may be available for segregation of the channel from the source (i.e. grouping based on temporal onset). Therefore, the extent to which the processes are separate is not entirely clear. Constancy across variation in pitch and dynamics appears to involve listeners’ ability to categorise sounds based on timbre despite changes in timbre that are caused by concurrent changes in another dimension of the sound (e.g. listeners may perceive timbre changes that are caused by changes in pitch but perceive the auditory object to be same because of their knowledge of pitch/timbre relationships). This again appears to be a different process to that of removing the channel, however, there may be overlap between this perceptual process and removing the effects of the channel/phonetic context. It will be seen in Chapter 5 that the appropriate categorisation of timbre despite changes in pitch/loudness and the perception of a sounds timbre relative to a change in
pitch/loudness that may responsible for timbral stability in this case, may also be involved in compensation for the channel.

1.4 Thesis overview

The previous section defined the scope of this thesis and some of the terms that will be used throughout. The current section illustrates the structure of the remainder of the thesis by giving an overview of each chapter.

Chapters 1, 2 and 3 address Research Question 1a. The primary aim of this question is to ascertain whether or not compensation for spectral colouration caused by loudspeaker and room acoustics occurs and the extent of this compensation.

Chapter 1: Introduced the role of timbre in recognising speech and other sounds, the problem of spectral distortion caused by the channel and the concept of timbral constancy and channel compensation.

Chapter 2: Will determine whether evidence from previous studies suggests that compensation for the spectral effects of loudspeakers and rooms occurs and discuss the best test methods to use to investigate this compensation further.

Chapter 3: Chapter 2 shows some evidence of compensation for loudspeakers and rooms when listening in conditions similar to those experienced in the real world. Chapter 3 will describe experiments that further confirm compensation for loudspeakers and rooms in real-world listening and measure its extent (Experiments 1 and 2). An experiment that begins to determine the mechanisms behind this compensation is also described (Experiment 3).

Chapter 3 shows that compensation for loudspeakers and rooms may partly involve a type of mechanism that is sensitive to time gaps between sounds. Chapters 4 to 6 of this thesis further investigate the potential mechanisms behind compensation for loudspeakers and rooms in real-world listening, with particular emphasis on determining mechanisms of compensation that are time-gap sensitive.
1.5 CHAPTER SUMMARY AND DISCUSSION

**Chapter 4:** Describes a research process for further determining potential mechanisms of compensation for loudspeakers and rooms in the real world when listening to speech and non-speech.

**Chapter 5:** Chapter 4 acknowledges that a literature review is necessary to further determine mechanisms of compensation. Chapter 5 describes a literature review of potential mechanisms of compensation for the channel in real-world listening focusing on mechanisms which might be time-gap sensitive. Most of this previous research is in speech perception and concerns compensation for a specific channel, the Vocal Tract (VT). Compensation for the phonetic context also occurs in speech perception and appears to involve a similar process to compensation for the VT. This review primarily aims to determine the extent to which mechanisms of compensation for the VT and phonetic context are potential mechanisms of the compensation for loudspeakers and rooms seen in real-world listening when listening to speech and non-speech programme items.

**Chapter 6:** In Chapter 5 it is acknowledged that specific compensation mechanisms may provide compensation for loudspeakers and rooms when listening to speech, but there is inadequate evidence that these occur when listening to non-speech sounds. Chapter 6 describes experiments which aim to determine the extent to which these specific compensation mechanisms cause compensation when listening with non-speech, and whether they can be considered potential mechanisms of real-world compensation for loudspeakers and rooms when listening non-speech, including music.

**Chapter 7:** Provides a summary and conclusion of the work described in this thesis and describes future work. This includes research to further answer Question 1c, thereby determining whether the potential mechanisms of compensation identified in this thesis actually produce compensation in real-world listening.

1.5 Chapter summary and discussion

This chapter introduced the research problem and set out a main research question and 3 sub questions (1a, 1b and 1c). This thesis focuses on answering questions 1a and 1b.
Question 1c will be answered in future work. Research Question 1a asks whether compensation for spectral colouration caused by the transmission channels, specifically loudspeakers and rooms, occurs and to what extent. Research Question 1b asks what the potential mechanisms are behind this compensation. Chapter questions were also presented which were to be answered via the work presented in the remainder of the chapter. These questions aimed to further clarify the research problem and further define some of the terms used in this thesis. This work is summarised and discussed via the answers to the chapter questions below. The extent to which this work answers Research Questions 1a and 1b is discussed in the conclusion to this section in brief and in the conclusion to this thesis (see Chapter 7).

1.5.1 Chapter questions

In section 1.2 chapter questions were set out. These chapter questions are answered:

**Question 1.1** asked: ‘How is timbre used in sound recognition in speech and non-speech perception?’ This was answered in Sections 1.3.1, 1.3.2 and 1.3.3. Timbre is the perception of the character, quality or colour of a sound so it is a key cue to identifying and recognising a sound. In speech perception timbre allows us to recognise different speech sounds. A number of attributes of a complex sound determine its timbre. Timbre is largely determined by the steady-state spectral content of a sound but other factors, such as the rate of attack and decay of a sound and the presence of non-harmonic partials, also contribute to timbre. The spectrum and spectral resonance peaks within each sound known as formants, are particularly important to our ability to recognise different speech sounds. The vocal tract size and shape determines the spectrum of speech sounds and the size and shape of the instrument body determines the timbre of non-speech sounds. Timbre can also be used to identify the talker in speech perception (the general size and shape of the VT varies between talkers as well as between speech sounds and this variation may contribute to giving each talker a unique speaking-voice timbre) and the identity of other transmission channels, such as loudspeakers.

**Questions 1.2** asked: ‘What is the effect of the transmission channel on timbre and
sound recognition?’ and **Question 1.3** asked ‘What is timbral constancy and how does this occur to reduce the effect of the channel to allow for more accurate auditory object identification?’ These questions were answered in Sections 1.3.6 and 1.3.7. The importance of the sound’s spectrum to recognition suggests a potential problem caused by spectral distortion due to the transmission channel. It was shown that spectral colouration or distortion by the channel changes the spectrum of speech and other sounds. This distortion occurs due to differences in amplification and attenuation of energy along the frequency spectrum of the sound, caused by the unique pattern of channel resonances and reflections. It was noted that transmission channels such as loudspeakers and rooms colour the sounds passing through them in a similar way to the vocal tract or other instrument body. This has the potential to affect the listeners’ ability to correctly identify an object using its timbre.

**Question 1.3** asked what is timbral constancy and how does this occur to reduce the effect of the channel to allow for more accurate auditory object identification? It has been observed that colouration caused by transmission channels does not impede listeners’ ability to recognise sounds as the perception of timbre appears to remain stable. This timbral constancy is thought to be the result of one, or a number of, perceptual processes, which may include mechanisms that compensate for the transmission channel spectrum. Mechanisms that provide timbral constancy may originate peripherally or centrally and they may be physiological or cognitive. It was noted that, in certain circumstances, mechanisms appear to exist that compensate for colouration by the VT and the phonetic context in speech perception and these may be the same as, or similar to, mechanisms that compensate for other channels.

Timbral constancy has been considered by some researchers to be a broad phenomenon which describes timbral stability despite changes to the spectral envelope of the sound reaching the ear due to other competing sounds in the environment, or changes in pitch or loudness of the sound source. In the current research a more limited research definition of timbral constancy was chosen. Timbral constancy may be defined as the stability of timbre relative to changes in a sounds’ spectral envelope that occur simply due to the sound passing through different channels. However, it was also noted that in speech perception the phonetic context can colour speech in a similar way to the VT and other channels and may be compensated for in the same way as the VT.
and other channels. Therefore, compensation for colouration caused by the phonetic context should be investigated as means of compensation for loudspeakers and rooms. Therefore, the research in this thesis will only examine the constancy that occurs given changes in a sounds’ spectral envelope when the sound passes through different channels, or in the case of speech, when it is heard in different phonetic contexts. It will not examine constancy of timbre despite changes to a sounds spectrum due to changes in pitch, loudness or competing sounds. It was observed that timbral constancy and compensation is unlikely to completely remove colouration by the channel, as some perception of the channel is likely to be useful for identifying the channel or the talker that produced a speech sound.

1.6 Chapter conclusion

The work in this chapter introduced the research problem. Timbre perception is important to our ability to recognise a sound. It was shown that timbre is largely caused by the perception of the spectrum of a sound and that the spectrum can be distorted or coloured by transmission channels such as loudspeakers and rooms. A process of timbral constancy, which may involve listeners compensating for the effects of this colouration may exist to improve recognition of sounds across transmission channels. More research is required to investigate this, particularly in relation to the extent to which colouration caused by specific channels—loudspeakers and listening rooms—is perceived, as this is of interest to audio engineers and hearing scientists generally. In light of this research problem this thesis aims to answer a main research question and 3 sub questions. This thesis will confirm that real-world compensation for spectral colouration caused by loudspeakers and rooms occurs and measure its extent (Research Question 1a), it will aim to determine potential mechanisms behind this compensation (Research Question 1b) and future work will be described to determine the actual mechanisms behind this compensation (Research Question 1c).

Research Questions 1a and 1b have been set out but not answered in this chapter. However, in relation to Question 1a it is noted that researchers have hypothesised that timbral constancy occurs in perception. This implies that there may be a
mechanism that compensates for channel distortion. Little information was provided to answer Question 1b but it was noted that timbre is processed at multiple levels in the hearing system, so compensation could occur at a number of points along the stimulus-to-perception chain. There was a suggestion that real-world listening may bring about increased compensation compared to laboratory listening and that there may be something about real world listening that causes this.
Chapter 2

Compensation for loudspeaker and room timbre

The previous chapter set the context for this thesis. It described the problem of spectral distortion caused by the transmission channel and the possible existence of a perceptual mechanism that compensates for this. A main research question and two sub questions were set out. Research Question 1a aims to determine whether compensation for spectral distortion caused by the channel affects the perception of loudspeakers and listening rooms. Literature specifically examining perceptions of loudspeaker and listening room timbre largely comes from audio engineering studies. The current chapter therefore reviews these studies to determine the extent to which colouration by the loudspeaker and listening rooms is perceived, and whether compensation occurs. Question 1b asks about the potential mechanisms behind this compensation. The main aim of the studies presented is to determine whether compensation occurs rather than examine its mechanisms but any discussion of mechanisms within these studies is mentioned here. Another issue of interest is the best way to measure the effect of channel compensation on the perception of loudspeakers and rooms. This will be discussed with a view to determining an appropriate paradigm to investigate compensation in the experiments in this thesis.

In Section 1.3.6 in Chapter 1 it was shown that objective measurements predict a perceptual influence of the loudspeaker and room spectrum on reproduced sound.
Resonances caused by the materials of the loudspeaker are evident in uneven frequency responses in loudspeaker measurements and room reflections and room modes result in modifications to the spectral envelope of sounds as they travel though the room (see Toole (2006) for a review of the objective effects of rooms and loudspeakers on reproduced sound). An aim of sound reproduction is to preserve the original timbral balance of a performance or any adjustments made by the recording engineer. It therefore usually desirable that the effects of the transmission channel are not perceived. However, general listening tells us that we can hear some spectral differences between loudspeakers and rooms and studies that have examined the effect of spectral modifications caused by loudspeakers and rooms on a sounds’ timbre generally show that channel effects are perceived to a significant and often large degree (see Toole (2006) for a review). These studies suggest a lack of compensation for loudspeakers and rooms.

However, firstly, it is not surprising that some perception of the channel remains as compensation is unlikely to be complete. It is not expected that the channel will be removed from perception entirely, just that its effects may be reduced to the extent necessary to maintain reasonable timbral constancy. Secondly, it is not clear from these experiments that compensation does not occur because:

a) Researchers have not compared subjective assessments of the loudspeakers/rooms against objective measurements. If perceived changes between loudspeakers/rooms are not as large as the changes reported in acoustic measurements then this may be evidence for compensation and,

b) Most studies that measure loudspeaker and room timbre do not measure perception in situations were compensation may be expected to occur. In particular they do not measure listening over the longer time periods where compensation might be expected to occur. In laboratory studies listeners tend to hear samples of loudspeakers/rooms for milliseconds to seconds at a time. This is because they are encouraged to switch between stimuli at will and listeners often choose to make rapid switches between stimuli to focus on the differences between them. Sometimes switching is very rapid (milliseconds) and sometimes it occurs after multiple seconds but listening for multiple minutes is much more
rare. Any compensation that occurs with longer listening may therefore not have had sufficient time to begin under these conditions.

Only a few studies have investigated the extent to which loudspeakers and rooms are heard in situations where compensation is more likely to occur (Bech 1994, Gilford 1979, Olive and Martens 2007, Olive et al. 1994, 1995, Schuck et al. 1993). In these studies the listening scenario is more similar to real-world listening. Listening is measured over a longer time period and there are gaps between auditioning different stimuli. Based on this real-world research there appears to be some evidence for compensation for loudspeaker and room acoustics. Some of the studies that investigate loudspeaker and room compensation are described in the following sections and the overall evidence for compensation will be discussed.

Based on the issues described above, two chapter questions are also set out in order to guide the work within this chapter towards providing information to answer Research Questions 1a and 1b.

Question 2.1) What is the evidence for timbral constancy and compensation for the channel in real-world listening to loudspeakers and rooms from prior experiments?

Question 2.2) What is the best way of measuring real-world compensation for loudspeakers and rooms in an externally valid manner and how will compensation be measured in this thesis?

This chapter is divided into sections detailing the measures of timbre used in studies investigating loudspeaker and room perception (Section 2.1), specific studies which appear to show compensation for loudspeaker and/or rooms (Section 2.2) and a discussion of the best way to measure compensation for loudspeakers and rooms in future research (Section 2.3). The chapter summary and discussion is presented in Section 2.4 and the answers to the chapter questions are summarised here. The chapter conclusion is presented in Section 2.5 where the extent to which the research questions can be answered is discussed in brief and is also discussed in Chapter 7.
2.1 Assessing room and loudspeaker timbre

Before a review of compensation studies, an overview of the methods of measuring loudspeaker and room timbre is given. The extent of the effect of loudspeakers and rooms on timbre can be measured by asking listeners to rate the timbral fidelity of sounds produced through loudspeakers or rooms. Deviations from high fidelity are evidence that the loudspeaker or room is having a perceivable effect on the sound. Further, deviations resulting in low fidelity are usually associated with perceptions of low sound quality (Schuck et al. 1993). Therefore, sound quality is often used to measure the effect of the loudspeaker or room. However, modifications to timbre caused by the loudspeaker and room may be regarded as high quality if they ameliorate spectral distortions in the programme material/other transmission channels through which the sound has passed (Gilford 1979, Olive et al. 1995).

Similarly, preference can also be used to assess timbre changes caused by loudspeakers and rooms. Differences in preference ratings of a single sound between different loudspeaker models, or between different rooms, demonstrate an effect of loudspeaker/room on the sound. Many of the studies discussed in this section have used preference measures rather than timbral fidelity or quality to examine the effect of the loudspeaker/room because an aim of these studies is to obtain measures of preference for particular products for commercial purposes and because rating preference is thought to be an easier task for the listener than rating quality or fidelity. The listener is likely to have greater understanding of what they prefer compared to what is high quality or fidelity. Further, rating preference appears to be a more cognitively simple task which still gives a measure of channel effects: listeners must decide whether they like a sound but not also whether their opinions of the sound match some external notion of quality and, judging quality is likely to involve cognitive dissonance if personal preference does not match a notion of high quality but rating preference alone is less like to bring about this conflict (Festinger 1957). However, preference may be a noisier measure than fidelity or quality because the listener may find it difficult to maintain a stable notion of preference across the task and because individual listeners may differ more widely in what they prefer compared to what they consider high fidelity.
A further factor affecting all three measures as measures of the spectral effect of the channel is the fact that the listener may find it hard to separate timbral from spatial or temporal factors. The listener must be asked to give timbral quality, fidelity or preference ratings if the effect of the channel on the spectrum of the sound is to be judged rather than any spatial or temporal effects of the channel. In particular, in ratings of rooms, spatial and timbral effects may not be easily separated. It is often not clear whether the subjective rating scales used in tests of compensation reflect the timbral effects of the loudspeaker/room or the spatial effects. Further, in some studies listeners are not asked to give ratings for timbral preference but for general preference for the sounds heard, making it more likely that both spatial and timbral factors both contribute to room/loudspeaker ratings (Toole 2008), and that any compensation measured is compensation for the spatial effects rather than spectral effects of the channel. To attempt to minimise the influence of spatial impression all of the studies described in this section, except for Olive and Martens (2007), have used monophonic reproduction played through a single loudspeaker. This is thought to allow the listener to focus on timbre rather than other factors (Bech 1994).

2.2 Compensation for loudspeaker and room acoustics

This section reviews audio engineering studies that measure the extent to which the loudspeaker and/or listening room timbre is perceived when listening to a source reproduced through one or both of these channels. The facts of each study are presented followed by a discussion of that work. The overall case for compensation for the loudspeaker and room timbre is discussed in section 2.2.5.

2.2.1 Gilford (1979)

Gilford (1979) showed that the frequency spectra of pulsed-glide test tones recorded in an anechoic room or in a normally treated studio looked very similar, according to objective spectrum measurements, when re-recorded after playback in a normally
treated listening room. This was because the listening room, which was common to
both examples, had the overriding acoustical effect on the sound, diluting smaller
differences caused by the recording environment. However, when Gilford tested
the perceptual influence of the recording rooms using listening tests, he found that
the different studio environments in which the recordings were made were easy to
distinguish. The effect of the listening room was therefore diminished in the listener’s
perception compared to what would be expected from the acoustical measurements.
Gilford (1979) stated that the binaural mechanism is responsible for the ability to ‘hear
through’ the room and identify the recorded environment. He noted:

‘The listener’s own room has an influence on transmitted speech since it
adds coloration and other effects of reverberation in the same way as the
studio. However, any additions from this cause are reduced by the binaural
rejection mechanism and assume relatively less importance [...] the studio
effects are meshed with recording, which is made in mono, whereas those
of the listening room are not. When listening directly with two ears one is
provided with an automatic mechanism for partially rejecting sound other
than that coming from the direction of the source to which one is listening.
This is an evolutionary faculty possessed by all higher mammals. When it
is inhibited by only having one channel of information, one is conscious of
the reverberant sound and extraneous noise to such an extent that speech
loses its intelligibility and music its definition’.

The mechanisms behind this hypothesised binaural compensation effect were not
elaborated on further. However, Gilford acknowledges that compensation is not
complete. Some listening rooms created larger distortions that overrode studio
characteristics. Gilford (1979) further tested the effect of the listening room on
perception of studio acoustics by asking recording engineers to listen to speech, recorded
in six studios, in four different listening rooms, and rate each recording for preference.
Pairwise comparisons of all studios were made before moving on to a different listening
room. The effect of both the studio and the listening room was perceived to some
extent. Listening rooms showed compensatory effects for deficiencies in the recording
studio; a room with a long reverberation time in the bass supported sound recorded
with a heavy bass cut, and a very dead room favoured a studio with long bass
reverberation. It was also noted that longer reverberation times in listening rooms greatly increased the inconsistency of the subjects’ answers and ‘the reverberation time allowable in listening room is longer than that allowed in the studio’. Overall the results suggest that listening room acoustics are perceived alongside the effects of the recording environment but the recording studio appears to have had a predominant effect. This shows that reverberation, if not timbral colouration, has a larger effect on perception when it is produced by the studio and reproduced through a loudspeaker, than if it is produced by the room in which the listener is actually present and that some compensation for the room occurs.

**Gilford (1979) discussion**

According to Gilford, the results are evidence for a compensation or rejection mechanism for the spectral or temporal effects of room reflections. He appears to suggest that uncorrelated sounds resulting from room reflections are instantly and automatically rejected in favour of more correlated monophonic direct sound (see also (Brandewie and Zahorik 2010, 2011, Watkins 1999, Zurek 1979)). The binaural rejection process hypothesised by Gilford appears to be similar to that involved in the instant or near instant suppression of the localisation effects of reflections involved in the precedence effect (Litovsky et al. 1999) and the instant or near instant suppression of comb filtering (Zurek 1979). It appears that because of its instant time course such a binaural mechanism would compensate for the room immediately. It does not take time to occur and would be expected to work fully in all the judgements made in Gilford’s studies. The fact that the room effect is still present in these judgements shows that this mechanism may not offer *complete* or even a large compensation for the channel and there is still a potential for other compensation effects to contribute to this. Further, the binaural mechanism suggested by Gilford appears be more useful for removing the spectral (and temporal) effects of room reverberation which arrive from different directions, but it does not appear to be useful for removing the spectral effects of other transmission channels, such as that of the loudspeaker which is imprinted onto the direct sound as well as its reflections. Other compensation mechanisms may be involved in the compensation for these channels.
The time course of listening may have also played some role in the compensation seen in Gilford’s work. Immediate comparisons between studios were made in a single listening room and significant time was spent in the same listening room. Less time was spent listening to each studio. A gradual reduction in the perception of sounds that are continuous or repeated could explain why the listening room effect was not as prominent as objective measures would suggest and not as prominent as the studio. Another explanation could arise from the fact that side-by-side comparisons, without time gaps, were available for studios but not listening rooms. There would have been a time gap between sessions within each listening room due to the time required to move between them. The studio comparisons might have benefited from spectral contrast effects due to being heard in close proximity which may have exaggerated timbral differences between them.

Gilford’s results indicate that some compensation for listening room reflections occurs in listening. It is not clear whether this compensation is largely for the spectrum of room reflections or the temporal effects of reflections, as the study did not specifically aim to measure room timbre. It is also not clear whether this occurred instantly or over the time spent in the room. The method of compensation for the listening room was proposed to be an instantaneous rejection mechanism but compensation may also be due to the longer listening periods within the listening room compared to the time spent listening to studio effects. It appears that whatever the case, compensation for the listening room occurred but was not complete. The room continued to influence sound preference, alongside studio effects. This shows that there may be a relative constancy provided by a room compensation process but listeners are likely to maintain an ability to perceive the size, shape and timbre of a space to some extent.

2.2.2 Olive et al. (1994)

A study by Olive et al. (1994) demonstrated that listening room acoustics had a significant effect on loudspeaker preference. Room acoustics were varied by placing the same loudspeaker in 4 different positions within a single L-shaped listening room, while keeping the listener position fixed. Each position within the room exhibited different acoustic characteristics. Variations in loudspeaker preference with position
were statistically significant \((p < .001)\) and larger than the variation across loudspeaker models \((p = .950)\). There was also a loudspeaker by position interaction \((p = .018)\), confirming that loudspeaker preference can vary significantly with position in the room. Position and programme material balancing effects similar to those found in Gilford were seen; a programme by position interaction \((p = .020)\) revealed that a more reflective position in the room was preferred for choral music compared to popular music. These interaction effects were only present when the study was carried out using binaural recordings of the loudspeakers in rooms. They were not present with live listening.

**Olive et al. (1994) discussion**

In this study, changes in position of the loudspeaker might have created changes in loudspeaker timbre by placing loudspeakers in different room modes or increasing the level of reflections with certain spectral characteristics. These timbral effects may have been perceived. However, the effect of the room may also be to do with variations in spatial impression/envelopment with loudspeaker position, rather than timbral variation. This finding does not necessarily show that room timbre is perceived. The reported effect of room position within this study could be to do with the order in which comparisons were made. Listeners were predominantly comparing locations, rather than loudspeaker models, and switched quickly between hearing the same loudspeaker in different locations, using a rotating chair. If compensation takes more than a few seconds, then this quick switching between positions may have meant that listeners did not have time compensate for the location but they may have had time to compensate for the loudspeaker which was heard for longer at any one time (see also Olive et al. (1995), summarised below). Hearing the within-room locations side-by-side may have enhanced timbral differences. Further, tests using binaural recordings, whereby the listener can make even more rapid switches between loudspeaker positions and loudspeaker models, showed more interactions compared to live listening tests. This may indicate that the quicker comparisons made through the use of these recordings, means that effects are more apparent compared to when longer listening periods inherent in live experiments are used. This is further evidence for an
absence of compensation with short listening time courses but increased compensation occurring when listening is over a longer time course.

### 2.2.3 Olive et al. (1995)

In a second study by Olive and colleagues (Olive et al. 1995) (also reported by Schuck et al. (1993)), compensation for loudspeaker characteristics was observed as well as compensation for the listening room. Three loudspeaker models were compared for preference. Listeners were asked to switch between the different loudspeaker models, listening to each within a single room. The experiment was conducted using binaural recordings rather than live listening. This meant that the listener was not physically within the listening room, listening to each loudspeaker, but they were played recordings of music over headphones that had been played through each loudspeaker and in the same listening room. This method allowed for blind comparisons of loudspeakers and quick switching between loudspeakers. After the listener had rated all 3 loudspeakers with one listening room, the test was then repeated 3 more times using the same loudspeakers but each time these were heard as if being played in a different listening room (i.e. each of the 3 loudspeakers were rated using recordings made in a new listening room until ratings had been made for 4 different rooms). An approximately 10 minute gap occurred between conducting the test using a new room. The effect of the loudspeaker on preference was measured by comparing preference ratings for each of the 3 loudspeakers averaged across the listening rooms. Room preference was also measured by considering how the average rating across all 3 loudspeakers differed between the rooms. The experiment showed a large effect of loudspeaker model on preference but a small (non-significant) effect of the room. The authors claimed that this was because listeners could compare the loudspeakers more directly by switching between them instantaneously to hear each loudspeaker side-by-side with the others but the rooms were only compared indirectly through changes in room across the test. The 10 minute break period between listening to recordings made in a new room meant that listeners could not hear the rooms presented side-by-side, as they could for loudspeakers. This may have resulted in listeners being less sensitive to the timbral differences between them. Additionally, listeners
spent longer periods continuously listening to sounds recorded within each room than they spent listening to each loudspeaker (they only listened to each loudspeaker for milliseconds or seconds before switching to another, but spent seconds to minutes listening to sounds recorded within a single room). It is possible that the longer listening to the room resulted in more compensation for the room compared to the loudspeakers. It appears that both longer gaps between listening to different rooms and longer continuous listening to each room may have caused smaller perceived differences between rooms (compensation for the effect of room). This did not occur for the loudspeakers because of the shorter listening time course.

Compensation with longer listening time courses was confirmed when the study was conducted again with the same rooms and loudspeakers but the order in which comparisons were made was changed (as before, binaural recordings were used rather than live listening). This time the listeners’ primary task was to compare rooms for preference. They heard a single loudspeaker that had been recorded playing music in each of the different rooms and could switch between listening to the recordings within different rooms instantaneously to hear each room side-by-side with the others. They made room comparisons directly. When they had made a preference rating for each room for that loudspeaker there was an approximately 10 minute break and then they conducted the same test with recordings of the second loudspeaker model being played within each of the same 4 rooms. They then repeated this for the 3rd loudspeaker. Loudspeakers were therefore now compared indirectly across the test, like the rooms had been in the first experiment. This time results showed a large effect of the room on preference and a small (non-significant) effect of the loudspeaker. There were also significant interactions between loudspeakers and rooms and between a number of other variables (programme type, within room position) in each experiment.

**Olive et al. (1995) discussion**

Olive et al.’s (1995) results appear to show that the perceived magnitude of timbral differences between loudspeakers and rooms varies depending on the time period within which comparisons are made. Shorter listening time courses (i.e. short continuous listening, no or short breaks between comparisons) result in larger perceived
differences between loudspeakers and rooms. Longer listening time courses (longer continuous listening or longer breaks between comparisons) result in smaller differences between loudspeakers and rooms. This decrease in sensitivity to timbral differences indicates that something about the longer listening time course seen with indirect comparisons causes compensation for the effects of the loudspeakers/rooms. The authors hypothesised that the time gaps between comparisons may be particularly important to the results. They note that time gaps reduce memory for timbral differences between stimuli, which might mean smaller perceived timbral differences. The authors consider this process a form of adaptation to or compensation for the loudspeaker and room that is memory-based. This hypothesis might also explain the results in Gilford (1979) and Olive et al.’s (1994) studies, as in those studies compensation was also seen for the factor that was heard with longer time gaps between comparisons. However, indirect comparisons also involved longer continuous listening and length of listening may be behind the compensation seen in this test. A fatigue or adaptation to the room over a longer listening period, whereby the timbre of a sound tends to flatten with longer listening, may occur. This may be similar to loudness adaptation (Hood 1950), whereby the loudness of a pure tone or noise reduces over a period of seconds to minutes. However, the authors did not put forward a length of listening explanation for the compensation seen in this study. Alternatively, time gaps may be important but for instigating other compensation mechanisms that are not necessarily memory based. Memory and fatigue are not known mechanisms of channel compensation. Their role in compensation is speculative. Memory is not known to produce an effect that might explain reduced channel differences seen with indirect comparisons: memory loss with time gaps would be more likely to increase variation in loudspeaker/room ratings rather than make ratings more similar, and there is no known process of timbral fatigue whereby the timbre of a sound tends to flatten with continuous listening. There may instead be established specific compensation mechanisms that can explain the compensation seen with longer time gaps or longer listening periods seen in these real-world tests.

Further points to note in relation to this experiment are that room by loudspeaker interactions are also a measure of compensation for loudspeaker and room effects. Interactions indicate that both factors are still having an effect on ratings. The
presence of significant room by loudspeaker interactions in this experiment means that compensation for loudspeakers or rooms cannot be considered complete even though the main effect of loudspeaker and room reduced to non-significance. In fact a reduction in main effects can be caused by an increase in interactions, so it could be that the results of this study are due to the interpretation of the statistical data. This is further investigated in the studies reported in Chapter 3. The test was also conducted with a live in-situ listening method. Like in Olive’s 1994 study this test resulted in smaller interactions between loudspeakers and rooms (and other factors), compared to the test conducted using binaurally recorded stimuli played over headphones. The listening periods in the in-situ listening test were considerably longer than those used in the binaural test. The decrease in interactions in the in-situ test therefore provides support for the claim that increased compensation for the loudspeaker and room occurs with longer listening periods.

Finally, the results may be caused by the experiment method influencing participants’ use of the rating scale, rather than any perceived reduction in timbral differences between loudspeakers and rooms. The task set-up or task instructions may have encouraged listeners to focus on differences between the factor being compared immediately (e.g. loudspeakers) and reporting these differences in ratings, rather than the factor that changed more slowly across the test (e.g. rooms). This is discussed further in experiments reported in Chapter 3.

To summarise, the compensation for loudspeakers and rooms seen in this study may be to do with a decrease in sensitivity to loudspeaker/room acoustics because of a longer time course of listening—either because of longer continuous listening or longer time gaps between listening. This may be a result of memory, or a fatigue that results from longer continuous listening but memory does not appear to have the correct nature to explain this compensation and fatigue is not a known compensation mechanism. There may be specific known compensation mechanisms that can explain the reduction in channel effect with longer listening/longer time gaps. Alternatively, the results could be due to task factors, rather than channel compensation, such as the specific instructions given to the listeners. This compensation and its mechanisms are examined further in the experiments described in Chapter 3.
2.2.4 Olive and Martens (2007)

A study by Olive and Martens (2007) was conducted to examine the hypothesis that listeners compensate for room acoustics with increased time spent continuously listening within the room. Further, the authors sought to test whether or not this room compensation occurs when listening to multichannel loudspeakers. Compensation for loudspeakers was not of specific interest in this study and was not measured. Olive and Martens (2007) hypothesised that, within certain limits, more time spent in the room will produce more compensation (less room effect). This is unlike the hypothesis in Olive et al. (1995) that compensation occurred because of the time gaps between comparisons of different rooms (and memory limitations), but is a test of the longer listening hypothesis.

Olive and Martens’ (2007) experiment used binaural recordings of multichannel loudspeaker configurations within rooms to allow quick comparisons between rooms as in Olive et al. (1995). One group of listeners rated 4 multichannel configurations in the same room successively, and then moved onto rating the same 4 sets of loudspeakers in 3 more rooms (as in Olive et al. (1995), where rooms were compared indirectly). Another group of listeners rated the same 4 multichannel configurations in different rooms presented non-successively, so that ratings for each configuration in the same room were interrupted by ratings in other rooms. This is similar to the condition in Olive et al. (1995) where rooms were compared by immediately switching between them (directly). The first condition was designed to increase the amount of continuous time spent in the same room; the second condition was designed to reduce the time spent in the same room. There were no break phases between listening to loudspeakers or rooms in either condition. There was a significant effect of room in both conditions and there was no significant room by condition interaction, confirming that overall room effects were equally large in successive and intermixed conditions. The authors claim that this shows that room compensation did not occur. However, there was a small difference between successive and intermixed conditions for room 4, hinting at a slight tendency to perceive room differences more easily in the intermixed condition. It is plausible that compensation effects in this room would be larger than the others as this room was noticeably more reflective and compensation for room acoustics may
occur to a larger extent in more reflective rooms.

**Olive and Martens (2007) Discussion**

In Olive and Martens’ (2007) study the difference between rooms was similar using both comparison methods: successive or intermixed. This shows that compensation did not occur overall. The authors suggest that this result may have been because the time spent listening in each room (1-2 mins) in the intermixed condition was too long and compensation therefore occurred in both conditions. There was also an absence of break phases between listening in either condition. This result may suggest that break phases are important in causing the reduction in loudspeaker/room effects, as suggested in Olive et al.’s (1995) work.

The use of multichannel surround-sound instead of monophonic reproduction may explain the difference between the results of this study and that of Olive et al.’s 1995 study. It is possible that the difference in preference ratings between rooms in this study is caused by spatial effects rather than timbral cues. Therefore, the finding of no compensation in this study does not necessarily preclude compensation to the timbral aspects of rooms. It is not clear how timbral and spatial effects of the room can be adequately separated but the use of a mono direct source rather than a multichannel source is likely to simplify matters. A mono source should be used in further experiments which aim to investigate compensation for the timbral effects of the room.

**2.2.5 Summary and discussion of compensation for loudspeaker and room acoustics**

The research described in this section shows that timbral differences between loudspeakers and rooms are reported to be smaller in some listening situations. This suggests that the perception of the loudspeaker or room is diminished and has been compensated for in these situations. It appears that the timbral effect of the sound that is changing more frequently is heard to a greater extent than any sound, such
as a static channel, that is constant for longer. The effects seen in these studies also appear to show timbral constancy whereby the effect of the unchanging factor (the channel—e.g. the room) is diminished and therefore the same source (e.g. music) heard through different channels (different rooms) will sound more similar across all channels because the different channels are no longer providing contributions to the timbre of that source.

The circumstances under which this compensation occurs, other than in situations of reduced spectral change, are not yet clear but it appears that the time course of listening is relevant to compensation. A longer time course (comprising of longer continuous listening and/or long time gaps between stimuli) appears to bring about more compensation compared to short listening time courses (short continuous listening and/or listening with no or short gaps between different stimuli—as in running speech or music). Most studies that investigate timbral differences between loudspeakers and rooms focus on comparing stimuli when the stimuli are frequently changing. For example in loudspeaker tests the listeners make direct comparisons of loudspeakers. These studies show large loudspeaker and room effects. However, in studies which examine loudspeaker and room effects over longer listening time courses, which are more common in real-world listening (i.e. in studies that measure effects when compared indirectly), the loudspeaker and room differences are not perceived to the extent suggested by laboratory studies. To measure the perception of loudspeakers and rooms in a way that more accurately represents real-world listening, tests should involve listening over longer time courses where compensation for the channel may occur.

It was noted that only a few studies have aimed to measure compensation that may occur with longer listening periods. Most have aimed to determine the extent to which the room is compensated for. Room compensation may be instantaneous (via binaural rejection) or occur over a longer listening time course. Evidence of compensation over a longer listening course is mixed. Two studies provide evidence of compensation for the room with longer listening (Gilford 1979, Olive et al. 1995) but only one directly compares this with short listening confirming compensation over a longer time period (Olive et al. 1995). Olive and Martens (2007) fail to show compensation for the room with long listening, but task factors, such as the use of multichannel rather than monophonic listening, may be behind this. Further confirmation of compensation
for the room over longer time periods is needed. This is investigated in the research described in Chapter 3.

Olive’s 1995 study shows that the loudspeaker is compensated for in the same way as the room when conditions are equivalent. Again compensation for the loudspeaker occurs with a longer listening time course but not when the time course is short. The fact that compensation occurs for the loudspeaker as well as the room is a good indication that it is not a binaural rejection mechanism that is providing the compensation for the room (according to Gilford this would not remove the effect of loudspeaker colouresation, which is carried in the direct sound.) Therefore, another compensation effect is probably at work which explains both compensation to loudspeakers and rooms and depends on the time course of listening. However, Olive et al’s studies were the only studies to confirm compensation for the loudspeaker as well as the room, when comparing short and long listening periods. Further confirmation that compensation for loudspeakers as well as rooms occurs is desirable and will be conducted as part of the research described in Chapter 3.

2.3 Determining the best method of measuring compensation for loudspeakers and rooms

Whether or not compensation occurs for both loudspeakers and rooms it appears that a time course sensitive mechanism is involved. The operation of a time course sensitive mechanism must be measured by comparing perception at different time courses: before compensation is expected to occur and afterwards. Measurement at short time courses (where it is not expected to occur) offers an appropriate baseline measure, against which compensation over longer listening periods can be measured. Compensation occurring over real-world time periods cannot be measured by showing perception is reduced compared to what is expected according to objective measurements (as Gilford and Olive 1994 do), as this does not necessarily capture compensation that is adaptive and happens over a longer time course. Also measurement in this way may simply show limits of the hearing system rather than compensation.
The paradigm used in Olive (1995) provides an appropriate short time course baseline for measuring compensation. The factor that is suggested to have been compensated for can be compared both when listening over a long and short time course, in the same test where conditions are equivalent. This is the first test that offers this baseline. Olive’s 1995 paradigm is also useful because it provides simultaneous measurement of compensation for two factors so comparisons of the size of compensation for two different channels or sources under the same conditions can be made. A further advantage of Olive’s paradigm is that the factorial design requires listeners to share the rating scale between the two factors being heard at once, this is likely to be representative of listening in the real world where the listener must judge a sound against other sounds occurring at the same time and make space in their perceptual range for all simultaneous factors. This is unlike many psychoacoustic laboratory tests of compensation mechanisms where stimuli are frequently tested in a more isolated way. Tests before Olive et al.’s 1995 study involve some elements of this paradigm but this paradigm combines all of these advantages into a single method.

There are some negative aspects to this paradigm: it is possible that it may not be sensitive enough to measure very quick (e.g. milliseconds) or very slow (e.g. > tens of minutes) acting compensation mechanisms. It is also noted that the listening lengths are not tightly controlled as the listener creates the stimulus changes (however, this can be controlled without altering the paradigm if necessary). The time course of the factor being indirectly compared is somewhat dependent on the time course of the directly compared factor and this dependency may mean that it is not possible to test all the time course manipulations necessary with this paradigm. Importantly, this paradigm has not been confirmed to elicit compensation. It has only shown compensation for loudspeakers and rooms once. Compensation for rooms in Olive and Martens’ (2007) study which used a variation of this paradigm was not found. Difference in the stimuli (multichannel loudspeakers) and method may explain this but the compensation for rooms seen in Olive et al. (1995) may be an anomaly. The paradigm should be shown to elicit compensation again, and ideally with a number of different types of channels, if it is an appropriate paradigm to use to examine the mechanisms of compensation generally.

For the reasons given above it can be concluded that Olive’s paradigm:
2.3 DETERMINING THE BEST METHOD OF MEASURING COMPENSATION FOR LOUDSPEAKERS AND ROOMS

a) mimics real-world listening by allowing the measurement of channel effects over longer time courses.

b) provides an appropriate baseline against which to measure compensation (listening over a short time course for the same factor). This measure of change tests compensation that occurs with longer listening periods and avoids the need to make comparisons with objective measures.

c) can be used to measure the effect of listening to more than one factor at once, thus showing whether loudspeakers and rooms are both compensated for in equivalent circumstances.

d) involves a factorial design means that results will show how perception depends on another factor and how the listener adjusts their preference scale when they need to consider two factors simultaneously, which contributes to the similarity of this listening test scenario to listening in the real world.

These points mean that this paradigm is useful for testing real-world compensation further and may be useful for testing the mechanisms of this. However, it is possible that this test will not be sensitive enough to measure all compensation, there may be some limitations when it comes testing the mechanisms of compensation using this paradigm and it is also possible that the paradigm will not show the same compensation effects in the future when the paradigm is tested with different stimuli. The compensation as seen in Olive et al. (1995) needs to be confirmed for both loudspeakers and rooms with different stimuli to those used in the original test to ensure that compensation occurs.

**Measures of timbral effects**

The work in this thesis aims to measure compensation for the spectral effects of transmission channels. Most of the studies described have used timbral preference ratings to show that the listener perceived spectral differences between loudspeakers and rooms. Rating for timbral *preference* may be preferable to rating timbral *quality* or *fidelity* as this represents an easier task for the listener. However, it is acknowledged that the listener may find that their preference changes across the task as they gradually
develop a preference. Experience with the stimuli before the test may prevent listeners changing their notion of preference throughout the test but added experience via a practice session may familiarise listeners to the extent that compensation is created in the familiarisation phase. If preference ratings are collected across many listeners there will be differences in preference between listeners. This may create overall noise in the measurements but should not affect the ability to measure compensation because it is only necessary to look at the overall preference rating across listeners and how difference in preference reduces with the time course of listening. This issue is less of a problem with quality and fidelity as there will be less variation between listeners in respect of this. Rating fidelity is not appropriate in tasks like Olive’s because there is no live performance to compare the sound against, so true fidelity ratings are not being made. Asking for ratings of quality may result in similar problems as with preference but also increase the complexity of the task by asking listeners to decide based on a complex notion of quality. It is concluded that preference is a suitable way of measuring perceived loudspeaker and room effects and the compensation for these. Like in Olive et al.’s (1995) experiment preference ratings will be used in future tests. An issue with all measures is that of ensuring that listeners judge the spectrum of the sounds they are listening to and make timbral ratings rather than spatial ratings. To avoid this listeners should be specifically asked to make timbral preference ratings. However, this may be difficult in the case of rooms where timbre varies less than spatial characteristics. Like in the studies in this chapter (expect Olive and Marten’s 2007 study) monophonic reproduction of the source will be used to aid listeners in rating timbre rather than spatial factors.

2.4 Chapter summary and discussion

The current chapter examined evidence for compensation for spectral colouration caused by loudspeakers and listening rooms obtained from audio engineering studies. An appropriate paradigm and measures to further investigate this compensation were discussed. Chapter questions were presented which were to be answered via the work presented. These questions aimed to guide the work towards providing information necessary to answer Research Questions 1a and 1b. The findings within this chapter
are therefore summarised and discussed via the answers to the chapter questions below. It was noted that the studies described in this chapter primarily provide information to answer Research Question 1a but that evidence regarding Research Question 1b would also be discussed. The extent to which Research Questions 1a and 1b can be answered from this work is discussed in brief in the conclusion to this section and in the conclusion to this thesis (see Chapter 7).

### 2.4.1 Chapter questions

The answers to the chapter questions are addressed:

**Question 2.1** asked: ‘what is the evidence for timbral constancy and compensation for the channel in real-world listening to loudspeakers and rooms from prior experiments?’ This was answered in Section 2.2. A number of studies have investigated loudspeaker and room perception. These studies show a significant effect of the loudspeaker/room on the timbre of a sound source. However, there is also evidence for compensation for the timbral effects of both loudspeakers and rooms when listening to music and other material reproduced through these channels. Timbral constancy is achieved by such compensation as this reduces the effects of the channel on the source. Compensation may occur instantly via a binaural rejection mechanism but the contribution of an instant mechanism appears small as loudspeaker and room effects are perceived to a large extent in laboratory listening. Loudspeaker and room effects appear further diminished with a longer listening time course compared to that used in laboratory listening. A longer listening time course may cause extra compensation in real-world listening but there may also be extra compensation caused by other differences between laboratory and real-world listening. One study by (Olive et al. 1995) has shown this extra compensation when using a paradigm that measures listening in a real-world like scenario (indirect comparisons of stimuli) and a laboratory like scenario (direct comparisons). This study used an appropriate paradigm to measure the extra compensation that occurs with real world listening as it directly compared listening over longer (real-world listening) and shorter (laboratory listening) time courses and used a factorial design. However, it was concluded that this real-world compensation needs confirming as this was the only study to show this. It was concluded that more
work is necessary to confirm the extent of compensation for loudspeakers and rooms in real-world listening, using this paradigm. Compensation for loudspeakers and rooms over real-world listening is unlikely to be complete.

**Question 2.2** asked: ‘what is the best way of measuring real-world compensation for loudspeakers and rooms in an externally valid manner and how will compensation be measured in this thesis?’ This question was answered in Section 2.3. Olive et al’s paradigm provides a good means for comparing listening over the time courses (and other circumstances) representative of real-world listening and laboratory listening via providing an appropriate baseline measure. Compensation is measured in a real-world listening situation compared to a laboratory situation and the extra compensation with real-world listening is measured. This paradigm includes both measuring the effect of longer continuous listening and longer time gaps between listening, both of which may cause compensation to channels in the real world. This comparison of short time course or *direct* and longer time course or *indirect* listening methods will be examined for further evidence of real-world compensation and the mechanisms behind this in this thesis (see Chapter 3). This method also allows comparison of compensation for different channels within the same test. The factorial design means that the listener must consider each factor in light of the other when making ratings and this is similar to real-world listening in light of multiple simultaneous factors (laboratory listening often requires listening to a factor interest in isolation). This style of listening may contribute to the real-world compensation seen in Olive et al’s studies. This paradigm is preferable to examining perception against objective measures (which would not measure *real-world* compensation over longer time courses and may not measure compensation at all). Timbral preference measures will be used to measure the extent to which loudspeaker and room effects are perceived. Adjustments to this paradigm and measures may be needed in future experiments because it may not be possible to measure very short or very long compensation using this paradigm or fully control the time course of listening in all circumstances.
2.5 Chapter conclusion

The work in this chapter contributed to answering Research Question 1a. Prior work shows that compensation for the spectral colouration caused by loudspeakers and rooms appears to occur. It is hypothesised that there is an instant compensation mechanism. This has not been confirmed and if it occurs it is not complete as the loudspeaker and room still have a significant effect on timbre in laboratory tests despite any such compensation. However, the fact that any instant mechanism does not remove compensation completely provides scope for additional compensation mechanisms. A mechanism that causes compensation in real-world listening scenarios was demonstrated. This compensation may occur due to the longer time periods involved in real-world listening. Real-world compensation is of primary interest to this thesis. It is noted that real-world compensation is also unlikely to provide complete channel compensation as some perception of the channel may be useful for identifying the room in which one is in or talker producing a speech sound. It is noted that further research is necessary to confirm real-world compensation and its extent.

The research in this chapter did not aim to directly answer Research Question 1b but some information was presented that contributed to determining the mechanisms of real-world compensation. Increased compensation with real-world listening is suggestive of compensation being due to the longer listening time course involved in real-world listening. Olive et al. (1995) have hypothesised that the time gap between listening with their indirect comparison method (i.e. in real-world listening) may cause this compensation. Further, they suggested that memory may be behind this compensation. It was also noted by this author that indirect comparisons involved longer periods of continuous listening to an unchanging channel and a mechanism that is sensitive to longer listening periods might cause real-world compensation, such as a process of *timbral fatigue*. However, these hypotheses regarding mechanisms of real-world compensation are speculative. Timbral memory does not appear to have the nature necessary to explain compensation effects and no *timbral fatigue* process is known to occur as yet. Specific compensation mechanisms may exist to explain real-world compensation. Upon confirming real-world compensation, mechanisms should be investigated further.
Chapter 3

Further confirmation of channel compensation in real-world listening

In the previous chapter a literature review of audio engineering studies showed that only one experiment provided evidence of compensation for loudspeaker and room colouration over a real-world listening time course, comparing short and long listening periods (Olive et al. 1995). It was acknowledged that further experiments should be carried out to confirm this compensation and measure its extent. It was also noted that Olive et al.’s paradigm appears to be suitable for measuring the extent of, and mechanisms behind, real-world compensation but compensation should be confirmed with this particular paradigm if it is to be used in future work. Further, if channel compensation is a genuine auditory process that is not specific to the stimuli used in Olive et al.’s test then compensation should occur when different rooms and loudspeakers are used and ultimately when any channel is used.

This chapter describes Experiment 1 which was conducted to replicate Olive et al.’s original study closely but using different loudspeakers and rooms in order to confirm compensation and extend this to channels not used in the original study. Experiment 2 was also conducted to confirm compensation but aimed to measure its extent under conditions more conductive to compensation than those in Experiment 1.
Additionally, Experiment 2 allowed for a preliminary investigation into whether compensation is caused by longer continuous listening, as was suggested in Chapter 2. Experiment 3 was conducted to further determine the mechanisms behind real-world compensation. Olive et al.’s primary hypothesis regarding the mechanism behind real-world compensation was that compensation is caused by the time between hearing different loudspeakers/rooms. Such time gaps may cause compensation via a memory process or other time-gap sensitive process. Experiment 3 aims to determine whether a type of mechanism that is time-gap sensitive is behind compensation. Additionally, if channel compensation is a genuine auditory process then this should not be explained by a particular feature of Olive et al.’s test that does not occur in real-world listening (e.g. it should not occur because of the specific instructions to listeners or some element of the experiment set-up). Therefore, Experiment 3 also examined whether the compensation seen in Olive et al.’s paradigm is to do with the task instructions given to participants. These experiments provide information which contributes to answering Research Questions 1a and 1b. In order to guide the work in this chapter towards providing answers to these questions, chapter questions are also put forward:

Question 3.1: Can compensation for loudspeakers and rooms seen in the real-world experiments described in Chapter 2 be elicited again with new stimuli?

Question 3.2: Is this compensation an artefact of the experimental process or is it likely to be a genuine listening phenomenon?

Question 3.3: What type of perceptual mechanisms might be behind this compensation?

Experiment 1, which replicates the experiment by Olive et al. (1995) with different loudspeakers and rooms, is described in full in Subsections 3.2-3.6. Experiment 2, which measures the extent of compensation with conditions more conducive to compensation is the described in Subsections 3.7-3.12. Experiment 3, which aims to determine whether a time-gap sensitive type of mechanism, or an aspect of the experimental process (i.e. the task instructions) are involved in compensation, is described in Subsections 3.13-3.19. The answers to the chapter questions will be discussed in the summary and discussion section of this chapter (see Section 3.20) and the extent to which Research Questions 1a and 1b can be answered via the work in this chapter is
discussed in brief here and in Chapter 7.

3.1 Experiment 1: Replicating and extending

Olive et al., 1995

This experiment replicates that of Olive et al. (1995) and extends this to confirm that compensation for loudspeakers and rooms occurs when using different loudspeakers and rooms. It is expected that, as was seen in Olive et al.’s (1995) experiment, larger differences between loudspeakers and rooms will occur when they are compared directly over a short time course than when compared indirectly over a longer time course that is more similar to listening in the real world. A statistically significant effect of comparison method (direct versus indirect), with smaller loudspeaker and room differences with the indirect comparison method, is expected. Follow-up experiments will then aim to uncover the processes behind this compensation. In order to replicate Olive et al.’s (1995) experiment closely most of the features of the experiment were kept the same. This experiment uses a similar group of experienced listeners (Section 3.2.1), the loudspeaker and room preference measures were kept the same (Section 3.2.2) and stimuli were produced using the same binaural reproduction method. However, the experiment extends that of Olive et al. (1995) by testing for compensation using different loudspeakers and rooms (Section 3.2.3).

The instructions to participants (Section 3.2.5) and experimental procedure (Section 3.3) closely replicate those used in Olive et al.’s (1995) study, however, the length of listening and the time between comparisons were reduced to allow for a more simple and shorter task, whilst aiming to still produce a significant compensation effect. The results were analysed using the same statistical methods as Olive et al. (1995). Additionally, the statistical significance of any compensation effects is reported. Analysis methods and results are reported in Section 3.4 and discussed in Section 3.5.
3.1.1 Research plan and hypotheses

Experiment 1 was designed to determine whether the findings in Olive et al.’s (1995) study could be replicated with different loudspeakers and rooms. The hypothesis (Hypothesis 1), tested in these experiments, consists of a number of more detailed sub-hypotheses (1a, 1b, 1c, 1d and 1e) to test for compensation. The sub-hypotheses are described in Table 3.1.

Hypothesis One

There will be compensation with indirect comparisons similar to that seen in Olive et al.’s (1995) study with different loudspeakers and rooms.
3.1 EXPERIMENT 1: REPLICATING AND EXTENDING OLIVE ET AL., 1995

Table 3.1: Experiment 1 hypotheses.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Olive et al. (1995)</th>
<th>This Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a) There will be a statistically significant main effect for both the room and loudspeaker factor, at the $p &lt; .05$ level when compared directly. There will be a non-significant main effect for both ($p \geq .05$) when compared indirectly.</td>
<td>For each factor (loudspeaker and room) Olive et al. (1995) conducted ANOVA analysis for each condition (direct and indirect comparisons) separately to test for significant differences. Larger differences were found when comparisons were direct and these were statistically significant at the $p &lt; .05$ level but when comparisons were indirect they were not ($p \geq .05$). Results can be seen in Olive et al. (1995) Appendices 2 and 4).</td>
<td>Olive et al. (1995) emphasised the importance of this pattern of statistical significance (significance with direct but not indirect comparisons) in demonstrating compensation. To exactly replicate Olive et al. (1995) significant differences in loudspeakers and rooms should be seen when comparisons are direct but not when they are indirect.</td>
</tr>
</tbody>
</table>

Continued on Next Page...
Table 3.1—Continued

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Olive et al. (1995)</th>
<th>This Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>1b) The loudspeakers and rooms used in this study are different to those</td>
<td>N/A</td>
<td>The range of loudspeakers and rooms in this study will be compared with</td>
</tr>
<tr>
<td>in Olive et al. (1995) and the range of stimuli tested may be different.</td>
<td></td>
<td>Olive et al. (1995). If there is wider timbral variation in loudspeakers/</td>
</tr>
<tr>
<td>The pattern of significance described in hypothesis 1a is not expected</td>
<td></td>
<td>rooms used here it is expected that there may be a significant effect at</td>
</tr>
<tr>
<td>to occur.</td>
<td></td>
<td>the p &lt; .05 level for both factors in both conditions, even where the</td>
</tr>
<tr>
<td></td>
<td></td>
<td>perceived difference between stimuli has reduced in the indirect</td>
</tr>
<tr>
<td></td>
<td></td>
<td>comparison condition. Likewise, smaller timbral variation is expected to</td>
</tr>
<tr>
<td></td>
<td></td>
<td>result in non-significant effects in both conditions.</td>
</tr>
</tbody>
</table>

Continued on Next Page...
<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Olive et al. (1995)</th>
<th>This Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>1c) There will be a significantly larger main effect for each factor when compared directly than when compared indirectly.</td>
<td>The $p$ value assessment used by Olive et al. (1995) does not test whether the difference caused by comparison method is large enough to be statistically significant. A change in main effect from $p &lt; .05$ in the direct condition to $p &lt; .06$ in the indirect condition is unlikely to be evidence of a statistically significant reduction in perceived variation with comparison method, but would meet the requirements in Olive et al. (1995) for compensation. As mentioned in hypothesis 1b, if the $p$ values show a significant effect for a factor in both conditions, but there is a significant reduction in effect between conditions, Olive’s method would lead to an incorrect rejection of the compensation hypothesis.</td>
<td>A statistically significant reduction in main effect is expected to occur with indirect comparisons. This will be tested by examining the main effect<em>condition interaction for each factor. It is expected that if compensation occurs there will be a significant main effect</em>condition interaction and a larger main effect in the direct comparison condition. These interactions are expected to be ordinal showing an expansion/contraction of preference ratings with comparison method but no changes in the order of preference between conditions.</td>
</tr>
</tbody>
</table>
### Table 3.1—Continued

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Olive et al. (1995)</th>
<th>This Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>1d) Room*Loudspeaker interactions are expected to be small or non-significant if compensation occurs.</td>
<td>Olive et al. (1995) reported significant room*loudspeaker interactions, but did not interpret these in light of their compensation hypothesis.</td>
<td>It has been reported in a number of tests that the room in which a loudspeaker is played affects loudspeaker ratings and vice versa. However, the compensation hypothesis assumes that where a factor is compensated for (i.e. when indirectly compared) the influence of that factor on ratings decreases. Room<em>Loudspeaker interactions should therefore be reduced or non-existent if compensation for the indirect factor occurs. Any significant loudspeaker</em>room interactions are evidence that compensation is not complete and that the factor undergoing compensation (the indirectly compared factor) still affects perception. Significant interactions will be noted and discussed in light of the compensation hypothesis.</td>
</tr>
</tbody>
</table>
3.2 Experiment 1: Overview and materials

A summary of the procedure for Experiment 1 is given here. A full description can be found in Section 3.3. This experiment followed the procedure reported in Olive et al.’s (1995) paper. Like in that experiment listeners were asked to compare and rate 3 different loudspeakers and 3 different rooms for timbral preference. However, the test was simplified compared to that used in Olive et al.’s work: loudspeakers and rooms were only tested for a single programme item and for a single within-room position, whereas 3 programme items and 3 room positions were tested in the original study (see Olive et al. (1995) for a full description of the original procedure and materials). This modification was made to reduce the length of the experiment whilst maintaining a paradigm which measured compensation over short and long listening time courses. Based on informal pilot experiments it was predicted that this simplification would not prevent the replication of compensation effects. As in Olive et al.’s (1995) study participants took part in 2 conditions: a loudspeaker rating condition (or a loudspeaker condition) and a room rating condition (or a room condition) where rooms were rated directly.

The loudspeaker condition

In the loudspeaker condition 3 different loudspeaker models (LS1, LS2 and LS3) were compared directly, within a single room (e.g R1). Comparisons between loudspeakers could be made by listening to each loudspeaker side-by-side by switching between them without a time gap between presentations. Listeners made as many switches between loudspeakers as they required. Instant switching was enabled via the use of binaural recordings of the loudspeakers within the room, played over headphones. Listeners listened to the loudspeakers for milliseconds or seconds before switching to another to make a comparison. The binaural recording listening method also ensured that listening was blind. Preference ratings were made for each of the 3 loudspeakers in that room, then the listener moved on to complete the same task in two more rooms (e.g R2 then R3). There was a 2 minute gap between listening in each new room. In the loudspeaker condition room ratings were obtained, indirectly, by examining how the overall ratings for the loudspeakers changed between rooms. The listener could also be
said to be making indirect room comparisons by acknowledging the room changes by changing their preference ratings between rooms. This method of indirect comparison involved the stimuli being rated when they are heard over a longer continuous listening period, and when there was time between listening (there were 2 minute time gaps between performing the loudspeaker rating task in each different room).

**The room condition**

The same loudspeakers (LS1, LS2 and LS3) and rooms (R1, R2 and R3) were used as in the loudspeaker condition. In the room condition the three rooms were compared directly for a single loudspeaker (e.g. LS1). The listener switched immediately between rooms, making preference ratings for each when ready. This was then repeated for each of the remaining two loudspeakers (e.g. LS2 then LS3) with a gap of 2 minutes between testing with each new loudspeaker. Loudspeakers were therefore compared indirectly.

Each condition followed a $3 \times 3$ factorial design and the listeners completed ratings in both the loudspeaker and room conditions in a counterbalanced order with a 24 hr or longer gap between conditions. After all participants completed both conditions the difference in preference between the 3 loudspeakers and the difference in preference between the 3 rooms was analysed separately for each condition: Loudspeaker and Room. The size of the difference for each factor was then compared between conditions to determine the effect of condition (i.e. the effect of making ratings directly or indirectly) on the size of the loudspeaker/room differences.

### 3.2.1 Participants

Eight students of the Institute of Sound Recording, University of Surrey participated in the experiment. All students had prior experience of psychoacoustic listening tests and most had undergone a technical ear training course. The students were aged between 20 and 35 and reported no hearing deficiencies.
3.2 EXPERIMENT 1: OVERVIEW AND MATERIALS

3.2.2 Rating scales

Participants were asked to rate each loudspeaker and room combination for timbral preference. The rating scale used was the same as that used in Olive et al.’s (1995) original experiment and derives from an early experiment by Olive and colleagues (Olive et al. 1994). A 100 point rating scale was used with the descriptors displayed at: 10 points, ‘really dislike’; 30 points, ‘moderately dislike’; 50 points, ‘neither like or dislike’; 70 points, ‘moderately like’; and 90 points, ‘really like’. Participants were told that they should rate stimuli that they ‘neither liked or disliked’ in the middle of the scale (50 points). As per the instructions given in Olive et al.’s (1995) work participants were told that when they had a ‘strong’ or ‘large’ preference between two stimuli they should use at least a 20 point difference between ratings to mark this preference. When they had a ‘moderate’ preference, they should use at least a 10 point difference, when they had a ‘slight’ or ‘small’ preference, at least a 5 point difference.

3.2.3 Stimuli

The programme material, loudspeakers, rooms, and within-room loudspeaker placement are described in this subsection.

Programme material

In Olive et al. (1995) 3 vocal programme items were used in order to examine interactions between programme item, loudspeakers and rooms. This was not an aim in this experiment so, to reduce experiment complexity, a single vocal track was selected:

Jennifer Warnes, “Bird on a wire” from Famous Blue Raincoat, Cypress Records, CD 258418

This track was chosen because it was used in Olive et al.’s (1995) experiment and has wide spectral and dynamic range. In Olive et al.’s experiment it was reported that, of the 3 tracks in their experiment, this track was the most revealing of differences between loudspeakers. An 8 bar section of the track was taken from the first verse.
Table 3.2: Objective characteristics of loudspeakers used in Experiment 1.

<table>
<thead>
<tr>
<th></th>
<th>Loudspeaker 1</th>
<th>Loudspeaker 2</th>
<th>Loudspeaker 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type</strong></td>
<td>Double balanced passive radiator. Floor standing. Tweeter elevation 100 cm</td>
<td>Active near-field monitor. Mounted on stand. Tweeter elevation 130 cm</td>
<td>2-way Passive bookshelf speaker. Mounted on stand. Tweeter elevation 130 cm</td>
</tr>
<tr>
<td><strong>Bass/mid range driver</strong></td>
<td>125 W, Class D, ICE power 102 mm/4” concave diaphragm. 1 forward facing active driver and 2 90° facing passive radiator drivers</td>
<td>150 W RMS 200 mm</td>
<td>165 mm (6.5”)</td>
</tr>
<tr>
<td><strong>Tweeter</strong></td>
<td>125 W, Class D, ICE power 19 mm/(3/4)” coated fabric dome with acoustic lens to improve high frequency directivity</td>
<td>60 W RMS 25 mm with centre plug waveguide</td>
<td>25 mm (1”)</td>
</tr>
<tr>
<td><strong>Frequency response</strong></td>
<td>50-23,000 Hz (±3 dB)</td>
<td>60-40,000 Hz (±3 dB)</td>
<td>60-20,000 Hz (±3 dB)</td>
</tr>
</tbody>
</table>

The section was looped, joining at zero crossings and cross-fading was used to produce a smooth transition, without silence, between looped sections using Reaper v4.22. When listening to any particular room or loudspeaker the track looped continuously and switching between loudspeaker and rooms was instantaneous and resulted in the listener continuing to hear the same point of the looping track for the new loudspeaker/room.

**Loudspeakers**

The vocal track was heard through 3 different loudspeakers via binaural recordings (see Subsection 3.2.3). Loudspeakers were chosen on the basis they were representative of high quality consumer loudspeakers, like in Olive et al. (1995). The characteristics of each loudspeaker can be seen in Table 3.2.
3.2 EXPERIMENT 1: OVERVIEW AND MATERIALS

Rooms and loudspeaker placement

The listener heard the stimuli as if played in 3 rooms via the use of binaural recordings (see Section 3.2.3). Olive et al. (1995) chose rooms that were representative of the range of listening rooms found in home environments. This experiment aimed to replicate the acoustical variation in rooms used in Olive et al. (1995). An office (Room 1), a studio control room (Room 2), and an ITU-R BS.1116 listening room (Room 3) were used. Photographs of each room can be seen in Figures D.1 to D.3 in Appendix D.

Room 1

An office was used to make the first set of loudspeaker recordings (Figure 3.1). This room measures $4.68 \times 4.10 \times 2.50$ m ($w \times l \times h$). RT60 between 500Hz and 1kHz is 0.31 seconds. Wooden desks were positioned against all walls and large windows covered the right side wall. The floors were carpeted with thin carpet. There were wooden bookshelves on each of the walls and a floor standing bookcase and printer immediately behind the loudspeaker. Desk chairs were removed from the room. The loudspeaker/listener placement followed that in Olive et al. (1995) as closely as possible. In all rooms the loudspeaker was positioned 1.1 m from the back wall, facing the Head and Torso Simulator (HATS). HATS was positioned 2.9 m directly in front of the loudspeaker. The tweeter height was raised 1.3 m above ground for each loudspeaker. The ears of the HATS were elevated to 1.3 m so that they were in the approximate position of a seated listener.

Room 2

The second room was a studio control room at the University of Surrey ‘Studio 2 Control Room’ (Figure 3.2). The dimensions of this rooms are $6.40 \times 5.60 \times 2.60$ m. In this room the HATS was placed nearer to the left side wall due to the presence of a large mixing console on the right hand side of the room. The RT60 is 0.18 seconds.

Room 3

The final set of loudspeaker recordings were made in an ITU-R BS.1116 listening room (Figure 3.3). The dimensions of this room are $7.35 \times 5.33 \times 2.5$ m. This room had carpeted floor, a lay-in-grid tile absorbent ceiling, and full range acoustic absorber
Figure 3.1: Room 1. University of Surrey, Room 09BC03b.

Figure 3.2: Room 2. University of Surrey, Studio Control Room 2.

boxes on the walls. The RT60 for this room is 0.25 seconds.

**Binaural impulse response capture method**

Binaural recordings were made with the aim of reproducing the perception of in-situ listening to a musical programme item played through a particular loudspeaker in a particular room, over headphones. This method of listening was also necessary to
replicate Olive et al’s study and to allow for the immediate comparisons of rooms necessary for making direct, side-by-side, comparisons. It is well established that binaural recordings offer an experience that is perceptually similar to listening in-situ (Hegarty et al. 2007, Olive and Welti 2009, Postel et al. 2011). Olive et al. (1995) found that the relative position of loudspeaker and room timbral preference ratings were similar when using binaural recordings and in-situ listening. The use of binaural recordings was therefore not validated further in the present experiment.

A diagram of the binaural recording set-up can be seen in Figure B.1 in Appendix B. A Cortex Instruments Head and Torso Simulator (HATS) fitted with 1/2 inch MK231 condenser capsule microphones (sensitivity 50 mV/Pa) was used to make recordings. The HATS utilises silicone rubber ears in the shape of human pinna to produce the head related transfer functions of an idealised human listener. The recordings made by the HATS were filtered using a diffuse field filter and a high pass filter set at 0.7 Hz. 10 dB gain was applied to improve the signal to noise ratio. Adobe Audition (v2.0) was used to record room and loudspeaker responses. The Aurora Generate Sweep plugin for Adobe Audition (Farina 2000) produced a 20 Hz-20,000 kHz, 15 s exponential sine sweep (44.1 kHz, 32 bit float). The sine sweep was played through the loudspeaker at a level corresponding to 73 dBA pink noise measured by the HATS. Left and right channel recordings (16 bit, 44.1 kHz) were made of the loudspeaker output. Recordings
were checked for left/right balance and artifacts. Recordings made at the left and right ear were convolved separately with the inverse filter of the original sine sweep using the Aurora Convolve With Clipboard Process plugin for Adobe Audition (v4.2) to produce impulse responses for each ear. First Block Auto Arrange was selected to determine the amount of gain reduction needed to constrain the impulse responses at both ears to -6 dB, to prevent clipping when later convolved with the musical stimuli. Binaural tracks containing the left and right ear impulse responses for each loudspeaker/room combination were produced in Audacity v2.0 and trimmed to remove silence. The stereo 8 bar looping audio track was converted to mono using Reaper v4.22. The binaural impulse response for the loudspeaker/room combination was then convolved with the mono audio track in MatLab (R2011a) to produce a binaural reproduction of each loudspeaker in each room, playing the programme item.

Separate sine-sweep recordings were made of each of the 3 loudspeakers playing the sine-sweep in each of the 3 rooms. The loudspeaker was placed in the same position within each room as described above. Each of these recordings was converted into an impulse response and convolved with the programme item. The aim of this was to obtain stimuli that consisted of the programme item heard as if it was played through each of the loudspeakers in each of the rooms (when played over headphones), and stimuli that allowed the listener to switch between loudspeakers and rooms immediately in the test. The stimuli were loudness balanced using an A weighted LEQ function created in Matlab and equal loudness was confirmed via pilot listening.

### 3.2.4 User interface

An experiment interface was created using MaxMSP v4.6. This interface consisted of 9 experimental interface pages separated by a 2 minute break page. Figure C.1 in Appendix C shows page 1 of the experiment interface, which is identical in appearance to the following 8 pages). The patch controlled the randomisation of the 3 levels of the factor to be directly compared (e.g. the loudspeaker) to 3 buttons (A, B and C) on each page, and the assignment of the 3 levels of the indirect factor (e.g. the room) to each experimental page. The program stored data on loudspeaker/room ratings and the time taken to complete each experiment page. Stimuli were reproduced
through Sennheiser HD 600 headphones. No equalization was performed to eliminate
the headphone response. A Focusrite Virtual Reference Monitoring box was used to
amplify the headphone signal.

3.2.5 Instructions

Instructions were given to the participants at the beginning of the experiment (see
Appendix A). The instructions described the task and informed participants as to
how they should use the rating scale. To prevent possible bias listeners were told
only that they were to listen to different ‘versions’ of a musical stimulus. They were
not told that the differences between versions was due to the loudspeakers or rooms
changing. Stimuli were not labelled to prevent participants being aware of stimulus
identity and when stimulus repeats occurred. Because of the lack of information, it was
expected that most listeners would not be aware that they were listening to binaural
reproductions of loudspeakers in different rooms. Post experiment reports revealed
that some participants believed that they were comparing musical recordings with
various filters or codecs applied. A couple of listeners revealed that they thought they
were listening to loudspeakers in rooms but they were not able to tell whether the
loudspeaker or the room was changing on each trial. Participants were told that the
experiment would consist of 2 sessions ‘condition 1’ and ‘condition 2’ (i.e. the room and
loudspeaker conditions) and that conditions would be separated by 24 hrs or longer.

3.3 Experiment 1: Procedure

Familiarisation phase

A familiarisation stage was not reported in Olive et al. (1995) but is common to listening
tests and can increase the reliability of results, so a short familiarisation phase was used
here. A familiarisation stage consisting of 3 experimental pages of the experiment, with
break phases removed between pages, was conducted to allow participants to familiarise
themselves with the range of stimuli to be presented, the task procedure and use of the
rating scale.
3.3 EXPERIMENT 1: PROCEDURE

Test phase

This section gives a more detailed description of the procedure set out in brief in Section 3.2. As stated in the experiment overview, the procedure replicated that used in the experiment by Olive et al. (1995) as closely as possible except that it was simplified by the use of a single programme item and a single within-room loudspeaker position. This had the effect of reducing the amount of time listeners spent listening to the indirect factor continuously compared to in Olive et al.’s experiment. The break phases were also reduced to 2 minutes compared to the approximately 10 minutes reported by Olive et al. (1995) These alterations meant that the total experiment time was shorter for each participant than in Olive et al.’s (1995) experiment, and importantly the time course of listening where stimuli are compared indirectly was shorter.

As mentioned in the overview, two experimental conditions were completed in this test in counterbalanced order. These were named according to the stimuli being directly compared. In the Loudspeaker condition, 3 loudspeakers were compared directly within a single room and then these comparisons were repeated in 2 further rooms with a 2 minute break between rooms (rooms were indirectly compared). In the Room condition 3 rooms were compared directly for a single loudspeaker and these comparisons were repeated for 2 more loudspeakers (loudspeakers were indirectly compared). The same loudspeakers and rooms were used in both conditions. Table 3.3 and the description below describe the whole process for a participant who completed the loudspeaker condition followed by the room condition.

3.3.1 Loudspeaker condition

The participant was seated at a laptop computer and presented with a set of instructions (see Appendix A), and the initial experiment page displayed was on the computer (see Appendix C). On this page the 3 different loudspeakers were randomly assigned to buttons labelled A, B and C. The participant was not told which loudspeaker was assigned to which button. The participant was instructed to press the start button to hear the loudspeaker assigned to button A playing the programme material. Once the button was pressed the programme item looped continuously
### Experiment 1: Procedure

Table 3.3: Typical order of presentation of stimuli for a participant completing the Loudspeaker condition and Room condition 24 h later

<table>
<thead>
<tr>
<th>Loudspeaker Condition</th>
<th>Room Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Page Comparisons Rm</td>
<td>Page Comparisons LS</td>
</tr>
<tr>
<td>1 LS1, LS2, LS3 1</td>
<td>1 R1, R2, R3 2</td>
</tr>
<tr>
<td>2 minute break</td>
<td>2 minute break</td>
</tr>
<tr>
<td>2 LS2, LS3, LS1 2</td>
<td>2 R1, R2, R3 1</td>
</tr>
<tr>
<td>2 minute break</td>
<td>2 minute break</td>
</tr>
<tr>
<td>3 LS3, LS1, LS2 3</td>
<td>3 R2, R1, R3 2</td>
</tr>
<tr>
<td>2 minute break</td>
<td>2 minute break</td>
</tr>
<tr>
<td>4 LS2, LS3, LS1 1</td>
<td>4 R3, R1, R2 3</td>
</tr>
<tr>
<td>2 minute break</td>
<td>2 minute break</td>
</tr>
<tr>
<td>5 LS1, LS3, LS2 3</td>
<td>5 R3, R2, R1 1</td>
</tr>
<tr>
<td>2 minute break</td>
<td>2 minute break</td>
</tr>
<tr>
<td>6 LS2, LS3, LS1 2</td>
<td>6 R1, R3, R2 2</td>
</tr>
<tr>
<td>2 minute break</td>
<td>2 minute break</td>
</tr>
<tr>
<td>7 LS1, LS2, LS3 1</td>
<td>7 R1, R2, R3 1</td>
</tr>
<tr>
<td>2 minute break</td>
<td>2 minute break</td>
</tr>
<tr>
<td>8 LS2, LS3, LS1 3</td>
<td>8 R3, R2, R1 3</td>
</tr>
<tr>
<td>2 minute break</td>
<td>2 minute break</td>
</tr>
<tr>
<td>9 LS3, LS1, LS2 2</td>
<td>9 R3, R1, R2 3</td>
</tr>
</tbody>
</table>

without silence. The participant then switched between loudspeakers at will by pressing the buttons A, B and C. The participant was instructed to make a timbral preference rating for each loudspeaker. Participants could switch between loudspeakers as many times as they liked in any order, and could listen to each one for as long as they liked, before making preference ratings. A slider below each loudspeaker was used to select a rating (preference ratings could only be made while that loudspeaker was playing). When ratings for all 3 loudspeakers had been made a continue button was pressed to move on to the break period.

The break period consisted of a 2 minute time gap. During this time a countdown timer was displayed on the user interface. The participant left the room, walked around in a different acoustic environment including background noise. The experimenter called the participant back 15 seconds before the end of the break to ensure that they were ready for the next part of the experiment. After the break period a new experimental page was presented. The same loudspeakers were newly randomly assigned to buttons.
A, B and C. This time listeners heard the programme item as if being played through the same loudspeakers but in a different room. The room was one out of the 3 (R1, R2, R3) selected on a latin squares basis. As before, the participant toggled between loudspeakers to audition each loudspeaker and make ratings. Another 2 minute break followed, after which time the participant returned to rate the same loudspeakers in the 3rd room. This procedure continued until all 3 loudspeakers (LS1, LS2, LS3) were rated in each of the 3 rooms (R1, R2, R3), 3 times. This provided 3 repeated ratings for each LS/Room combination, for each participant. The participant returned at least 24 hours later to complete the room condition (a participant starting in the room condition returned at least 24 hours later to complete the loudspeaker condition).

### 3.3.2 Room condition

This condition follows the same procedure as the loudspeaker condition, except that on each experimental page the subject heard 1 loudspeaker (e.g LS1), playing in 3 different rooms (R1, R2, R3). The rooms were randomly assigned to the buttons A, B and C. Participants switched between the rooms to compare them and made ratings for preference when ready. When all 3 rooms were rated the listener pressed the continue button to move onto the break phase. After a 2 minute break (following the same procedure as for the loudspeaker condition), the participant rated the rooms for a different loudspeaker. This was followed by another 2 minute break then ratings of the same rooms for the final loudspeaker. This process continued until the participant had compared each of the 3 rooms (R1, R2, R3) for each of the 3 different loudspeakers (LS1, LS2, LS3), 3 times.

### 3.4 Experiment 1: Results

The data were analysed using repeated measures analysis of variance (RM ANOVA) in SPSS v 19. This method of analysis was used in Olive et al. (1995) to look for significant main effects for loudspeakers and rooms in each condition and room*loudspeaker interactions. In order to statistically test any compensation seen, the current study
also examines the moderating effect of condition (direct versus indirect comparisons) by examining loudspeaker*condition and room*condition interactions. Data from the familiarisation phases were not analysed.

The next subsection reports statistical analysis to check the reliability of measurements and to check that ANOVA requirements are met (Section 3.4.1), descriptive measures of loudspeaker and room compensation (Section 3.4.2) and statistical analysis to confirm the statistically significance of effects observed (Section 3.4.3).

3.4.1 Repeated measures ANOVA requirements and reliability

The residuals around the mean score for each loudspeaker/room combination were examined for normality using Q-Q plots. The data was normally distributed. Mauchley’s test of sphericity was run to ensure that data met the assumptions for repeated measures ANOVA analysis. All Mauchley’s tests were non-significant at the $p < .05$ level, indicating no significant departure from sphericity. Non-adjusted degrees of freedom were therefore used to compute variance measures.

Individuals’ scores were examined for intra-rater reliability. When making repeated ratings for the same stimulus in the same context, individuals’ scores were quite varied indicating low reliability within participants. Cronbach’s alpha was used to measure the reliability of each participants’ scores across repeats, and across all stimulus combinations (see Figures E.1 to E.8 in Appendix E). Cronbach’s alpha was below the acceptable level of .700 for 3 participants. Olive et al. (1995) did not report the removal of participants based on varied repeated ratings. For this reason the main analysis was conducted using all participants. Additional analysis with participants with low Cronbach’s alpha scores removed is reported separately.

3.4.2 Descriptive measures

This section examines graphs and descriptive statistics with the aim of assessing whether differences between loudspeakers and rooms are smaller when compared
indirectly, compared to directly and therefore whether compensation for the channel occurs.

**Loudspeaker factor**

Figure 3.4 shows the ratings for each loudspeaker in each room (averaged across all participants and repeated ratings). Figure 3.4 (i) displays for the loudspeaker condition (the direct comparison condition, for the loudspeaker) and the room condition (ii) (the indirect comparison condition for the loudspeaker factor). The method described in Loftus and Masson (1994) was used to adjust 95% confidence intervals for repeated measures analysis.

Figure 3.4 shows an increase in the spread of loudspeaker ratings within each room when comparisons are made directly, compared to indirectly. This effect can be seen most clearly when results are examined on a room by room basis. Loudspeakers are perceived to be more different from each other with direct comparisons compared to indirect comparisons for all rooms. Table 3.4 shows the variation in mean loudspeaker ratings within each room for both conditions. This table shows that the variance is larger in the direct comparison condition, compared with the indirect condition, for all rooms. Caution should be taken when comparing variance between conditions by eye as variance is measured in preference units squared. For this reason standard deviations are also shown. The ratio of standard deviations in the direct comparison condition to the indirect condition shows an increase in variance with direct comparisons for all rooms. These ratios show that there is moderately larger variation in the direct comparison condition and that the size of the effect of condition is slightly different for each room. The effect of condition appears to be largest for R1 ($3.81/1.79 = 2.13$) and smallest for R3 ($10.99/8.49 = 1.29$).

Figure 3.5 shows the main effect of loudspeaker in both conditions. The difference between loudspeakers is slightly larger in the condition where they are compared directly. This figure also shows that the difference between loudspeakers is mainly driven by the difference between LS1 and and the other two loudspeakers and the compensation effect is driven by the fact that this difference reduces between direct
and indirect comparison conditions. There is also a slight contraction for LS3. This reduction in main effect may indicate a statistically significant moderating effect of comparison condition (direct versus indirect), or the reduction in main effect may be too small to show significant compensation. Compensation is only expected to affect the degree of preference not the order of preference. Any change in order of preference with condition would not indicate compensation but some other effect of the change in comparison method. The order of preference in each room is the same across conditions in this experiment. This stability in ratings across conditions indicates that there is reliability in mean loudspeaker/room ratings across participants, that is demonstrated when the test is repeated by each participant 24 hours later, even though individuals’ ratings were sometimes unreliable.

### Loudspeaker factor summary

The data presented shows larger variation in preference between loudspeaker models in the direct comparison condition in all rooms. However, the standard deviation
Table 3.4: Loudspeaker mean ratings within each room and loudspeaker mean variance within each room.

<table>
<thead>
<tr>
<th>Loudspeaker</th>
<th>Condition</th>
<th>Room</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>R1</td>
<td>R2</td>
<td>R3</td>
<td></td>
</tr>
<tr>
<td>LS1 Mean</td>
<td>LS (Direct)</td>
<td>46.71</td>
<td>42.29</td>
<td>38.25</td>
<td></td>
</tr>
<tr>
<td>LS2 Mean</td>
<td></td>
<td>53.92</td>
<td>55.33</td>
<td>50.83</td>
<td></td>
</tr>
<tr>
<td>LS3 Mean</td>
<td></td>
<td>52.46</td>
<td>51.42</td>
<td>60.13</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>51.03</td>
<td>49.68</td>
<td>49.74</td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td></td>
<td>14.52</td>
<td>44.78</td>
<td>120.53</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td>3.81</td>
<td>6.69</td>
<td>10.99</td>
<td></td>
</tr>
<tr>
<td>Condition Mean</td>
<td></td>
<td>50.15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition Variance</td>
<td></td>
<td>59.94 (SD = 7.74)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD Ratio</td>
<td></td>
<td>2.13</td>
<td>1.54</td>
<td>1.29</td>
<td></td>
</tr>
</tbody>
</table>

ratios reveal that the difference in variation between conditions may only be considered moderate (up to a doubling of SD) and that the size of the effect of comparison method is different in each room. When ratings are averaged across all rooms a larger main effect of loudspeaker in the direct condition compared to the indirect condition can be seen. This difference may be large enough to result in a significant moderating effect of condition. The order of loudspeaker preference does not change between conditions as is expected by the compensation hypothesis.

Room factor

A similar compensation effect appears to occur when listeners are asked to rate rooms. Figure 3.6 shows the mean rating for each room for each loudspeaker, averaged across participants and repeats. This is displayed for each condition: the room condition (as shown in Figure 3.6,(i)) (this is the direct comparison condition for the room factor) and the loudspeaker condition (ii) (the indirect comparison condition). Participants had a slight tendency to rate the rooms as more different to each other when they are
Figure 3.5: Main effects for the loudspeaker factor: loudspeaker means averaged across all rooms for the Loudspeaker condition (i) and the Room condition (ii).

making direct room comparisons. However, this effect only appears to be strong for LS2. The effect is weaker for LS1 and LS3. The order of room preference changed slightly between conditions, which shows disordinal interactions that are not expected by the compensation hypothesis.

Table 3.5 shows the variation between rooms and standard deviations, for each loudspeaker, in each condition. The variation between rooms is larger for the direct comparison condition compared with the indirect comparison condition in all cases. Across condition variation between rooms is also larger for the direct comparison condition. However, the standard deviation ratios show that the reduction in variation with indirect comparisons conditions is small to moderate. The magnitude of reduction is different between loudspeakers. The room variance reduced least for LS1 ($SD_d/SD_i = 1.10$) and most for LS2 ($SD_d/SD_i = 2.18$).

Figure 3.7 shows the main effect of room in the room condition (i) and in the loudspeaker condition (ii). It can be seen that there may be a significant main effect of room in the direct comparison condition. It is unlikely that there is a main effect of room in the indirect condition. This may mean a significant main
### EXPERIMENT 1: RESULTS

**Figure 3.6:** i) Mean preference scores for each room in the Room condition. ii) Mean preference scores for each room in the Loudspeaker condition. Larger variation between rooms is expected in the room condition.

**Figure 3.7:** i) Main effects for the room factor in Room condition (direct comparisons). ii) Loudspeaker condition (indirect comparisons). Means are averaged across all loudspeakers.

Effect by condition (direct versus indirect) interaction in the direction suggestive of compensation. However, the difference between rooms in both conditions is small and the reduction in this difference between condition may be too small to show significant
Table 3.5: Room means within each loudspeaker and room variance within each loudspeaker.

<table>
<thead>
<tr>
<th>Room</th>
<th>Condition</th>
<th>Loudspeaker</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Room</td>
<td>LS1</td>
<td>LS2</td>
<td>LS3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Direct)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1 mean</td>
<td></td>
<td>45.79</td>
<td>49.33</td>
<td>48.00</td>
<td></td>
</tr>
<tr>
<td>R2 mean</td>
<td></td>
<td>51.79</td>
<td>59.33</td>
<td>51.79</td>
<td></td>
</tr>
<tr>
<td>R3 mean</td>
<td></td>
<td>42.63</td>
<td>53.79</td>
<td>59.29</td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td></td>
<td>46.74</td>
<td>54.15</td>
<td>53.03</td>
<td></td>
</tr>
<tr>
<td>variance</td>
<td></td>
<td>21.68</td>
<td>25.10</td>
<td>33.02</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td>4.66</td>
<td>5.01</td>
<td>5.75</td>
<td></td>
</tr>
<tr>
<td>R1 mean</td>
<td></td>
<td>46.71</td>
<td>53.92</td>
<td>52.46</td>
<td></td>
</tr>
<tr>
<td>R2 mean</td>
<td></td>
<td>42.29</td>
<td>55.33</td>
<td>51.42</td>
<td></td>
</tr>
<tr>
<td>R3 mean</td>
<td></td>
<td>38.25</td>
<td>50.83</td>
<td>60.13</td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td></td>
<td>42.42</td>
<td>53.36</td>
<td>54.67</td>
<td></td>
</tr>
<tr>
<td>variance</td>
<td></td>
<td>17.91</td>
<td>5.29</td>
<td>22.62</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td>4.23</td>
<td>2.30</td>
<td>4.76</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Condition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td>51.31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>variance</td>
<td>26.60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(SD = 5.16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SD Ratio</td>
<td>1.10</td>
<td>2.18</td>
<td>1.21</td>
<td></td>
</tr>
</tbody>
</table>

Room factor Summary

There appears to be a compensation effect whereby there is larger variation in rooms for all loudspeakers when direct comparisons are made compared to when indirect comparisons are made. However, the SD ratios show that the effect of comparison method is small for LS1 and LS3. Main effects analysis shows that the difference between rooms may be significant in the direct comparison condition but not in the indirect condition. Therefore, the expected compensation has occurred. However, the difference between rooms when compared directly may be too small to be significant in which case no further reduction indicative of compensation can be significant.
3.4.3 Statistical analysis

This section describes statistical analysis conducted to assess whether the effects described in the descriptive measures section are statistically significant and to test the specific hypotheses set out in Section 3.1.1.

Hypothesis 1a—Main effects Analysis

The significance of main effects in each condition was measured using separate ANOVA analysis for each condition. Results are presented in Table 3.6.
**Table 3.6:** Anova analysis: Main effects and loudspeaker*room interactions for each factor in the loudspeaker and room conditions.

<table>
<thead>
<tr>
<th>Factor</th>
<th><strong>Loudspeaker Condition</strong></th>
<th></th>
<th></th>
<th><strong>Room Condition</strong></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Type III sum of squares</strong></td>
<td><strong>Df</strong></td>
<td><strong>Mean Square</strong></td>
<td><strong>F Ratio</strong></td>
<td><strong>p value</strong></td>
<td><strong>Type III sum of squares</strong></td>
</tr>
<tr>
<td><strong>Loudspeaker</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error (Loudspeaker)</td>
<td>6517.148</td>
<td>2</td>
<td>3258.574</td>
<td>23.121</td>
<td>.000</td>
<td>2300.583</td>
</tr>
<tr>
<td></td>
<td>6483.074</td>
<td>46</td>
<td>140.936</td>
<td>1150.292</td>
<td>.000</td>
<td>1605.361</td>
</tr>
<tr>
<td><strong>Room</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error (Room)</td>
<td>83.676</td>
<td>2</td>
<td>41.838</td>
<td>0.291</td>
<td>.749</td>
<td>1605.361</td>
</tr>
<tr>
<td></td>
<td>6610.546</td>
<td>46</td>
<td>143.708</td>
<td>802.681</td>
<td>2.104</td>
<td>175551.08</td>
</tr>
<tr>
<td><strong>LS*Room</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error (LS*Room)</td>
<td>2115.102</td>
<td>4</td>
<td>528.775</td>
<td>4.232</td>
<td>.003</td>
<td>2224.806</td>
</tr>
<tr>
<td></td>
<td>11496.009</td>
<td>92</td>
<td>124.957</td>
<td>556.201</td>
<td>.000</td>
<td>83335.417</td>
</tr>
</tbody>
</table>
There was a significant main effect of loudspeaker in the loudspeaker condition, \(F(2, 46) = 23.12, p < .001\). There was also a significant effect of loudspeaker in the room condition \(F(2, 46) = 10.56, p < .001\). These results are contrary to Hypothesis 1a, which expects a significant main effect for loudspeakers in the loudspeaker condition only. There was no significant main effect of room in either the room condition: \(F(2, 46) = 2.10, p = .134\), or the loudspeaker condition \(F(2, 46) = 0.29, p = .749\). A non-significant main effect for room was only expected in the loudspeaker condition. Hypothesis 1a is therefore not supported for the loudspeaker factor or room factor. This analysis was also conducted with 3 participants removed from the analysis because their Cronbach’s alpha score was lower than .700 in the speaker condition. In this analysis effect of loudspeaker was significant \((p < .001)\) in the loudspeaker condition and in the room condition \((p = .005)\). The effect of room was not significant in the room condition \((p < .533)\) or in the loudspeaker condition \((p < .354)\). The removal of these participants does therefore not change the findings for Hypothesis 1a.

**Hypothesis 1b—Range of stimuli**

The fact that there is a significant effect for loudspeaker in both conditions may be because the range in loudspeakers chosen for this experiment is larger than in Olive et al.’s (1995) experiment, making significant main effects persist even when a reduction in variation occurs with indirect comparisons. Likewise, the variation in rooms chosen for this study might be smaller than in the previous experiment making non-significant effects persist even when direct comparisons are made. The variance for loudspeakers and rooms can be compared with that in Olive et al.’s (1995) work to determine whether there is a wider range of loudspeakers or a narrower range of rooms in this study. Only the direct comparison condition will be used in this analysis as this condition is expected to be almost identical across experiments, whereas the simplification of the method means that there are length of listening and time gap differences, which make indirect comparison conditions less similar. The relevant information about loudspeaker and room variation in Olive et al. (1995) can be found in Appendices 2 and 4 in Olive et al.’s (1995) work. These measures are reproduced here in Table 3.7 along with the corresponding information from the current study (this data is from the analysis
Table 3.7: Variance measures used to calculate main effect of loudspeakers and rooms when compared directly.

<table>
<thead>
<tr>
<th>Type III sum of squares</th>
<th>Loudspeaker factor (Loudspeaker condition)</th>
<th>Room factor (Room condition)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean Squares</td>
<td>$F$</td>
</tr>
<tr>
<td></td>
<td>Olive et al. (1995)</td>
<td>48.404</td>
</tr>
<tr>
<td></td>
<td>This Experiment</td>
<td>6517.148</td>
</tr>
<tr>
<td></td>
<td>Olive et al. (1995)</td>
<td>415.342</td>
</tr>
<tr>
<td></td>
<td>This Experiment</td>
<td>1605.361</td>
</tr>
</tbody>
</table>

presented in Table 3.6). Because of non-sphericity Huynh-Feldt adjusted $p$ values should be examined for Olive et al.’s (1995) data.

**Loudspeaker factor**

$F$ ratios and $p$ values in Table 3.7 show that there is larger perceived variation in loudspeakers for the direct comparison condition in this study than in Olive et al.’s (1995) work, after error is accounted for ($F = 23.121$, $p < .001$ in this study versus $F = 6.327$, $p < .017$ in Olive et al.’s (1995) study). This larger range in loudspeakers may explain why this factor continues to be significant when measured indirectly in this study.

**Room factor**

The initial variation in rooms is also larger in this study but there is more measurement error resulting in a smaller $F$ ratio for rooms than in Olive et al. (1995). It appears that participants were less reliable when rating rooms in this study than in Olive et al.’s (1995) experiment and larger error, rather than a smaller range in rooms explains why this factor is non-significant in both direct and indirect comparison conditions.

These issues explain why similar $p$ values to those in Olive et al.’s (1995) experiment
were not obtained and why Hypothesis 1a is not supported. Further analysis must be undertaken to determine whether or not there is a significant reduction in perceived loudspeaker/room variation with indirect comparisons.

**Hypothesis 1c—Effect of comparison method**

Although the pattern of significance expected by Hypothesis 1a was not found, it can be seen in Table 3.6 that $F$ ratios are smaller when indirect comparisons are made. It can be also be seen that the reduction in $F$ with indirect comparisons is more due to decreased variance for each factor rather than an increase in measurement error. This is expected by the compensation hypothesis. If the reduction in main effect with indirect compared to direct comparisons is statistically significant, this is evidence for loudspeaker and room compensation. A single $3 \times 3 \times 2$ (loudspeaker, room, condition) repeated measures ANOVA was used to test for a significant reduction in variance between direct and indirect comparison conditions for both loudspeaker and room ratings. A significant main effect*condition interaction, with larger variation in the direct comparison condition is expected, for each factor, if compensation has occurred. The results are shown in Table 3.8. The results for the room factor are presented for completeness but should not be further analysed as differences between room were not statistically significant to begin with.

**Loudspeaker factor**

The loudspeaker*condition interaction is not significant at the $p < .05$ level but is significant at the $p < .10$ level, $F(2,46) = 2.677$, $p = .079$. As this is a preliminary study, a higher probability of a Type 1 error is acceptable and effects significant at a $p < .10$ level are worthy of further investigation. When unreliable participants were removed this interaction increased to $p < .023$. However, Olive et al. did not remove unreliable participants and the removal of such participants reduces validity of the test. Therefore, only the main results should be considered. It is necessary to examine these statistics alongside the pattern of data shown in Figures 3.4 and 3.5 to confirm that the interaction demonstrates an effect that is in the direction suggestive of compensation. It was discussed in Section 3.4.2 that the data shows that loudspeaker variation reduces
Table 3.8: Main effect by condition interactions

<table>
<thead>
<tr>
<th></th>
<th>Type III sum of squares</th>
<th>Df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>LS</td>
<td>8171.477</td>
<td>2</td>
<td>4085.738</td>
<td>31.633</td>
<td>.000</td>
</tr>
<tr>
<td>Error (LS)</td>
<td>5941.301</td>
<td>46</td>
<td>129.159</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Room</td>
<td>497.977</td>
<td>2</td>
<td>248.988</td>
<td>1.037</td>
<td>.363</td>
</tr>
<tr>
<td>Error (Room)</td>
<td>11041.801</td>
<td>46</td>
<td>240.039</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>144.676</td>
<td>1</td>
<td>144.676</td>
<td>0.075</td>
<td>.786</td>
</tr>
<tr>
<td>Error (Condition)</td>
<td>44237.102</td>
<td>23</td>
<td>1923.352</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS*Room</td>
<td>4201.481</td>
<td>4</td>
<td>1050.370</td>
<td>10.466</td>
<td>.000</td>
</tr>
<tr>
<td>Error (LS*Room)</td>
<td>9232.741</td>
<td>92</td>
<td>100.356</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS*Condition</td>
<td>646.255</td>
<td>2</td>
<td>323.127</td>
<td>2.677</td>
<td>.079</td>
</tr>
<tr>
<td>Error (LS*Condition)</td>
<td>5552.968</td>
<td>46</td>
<td>120.717</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Room*Condition</td>
<td>1191.060</td>
<td>2</td>
<td>595.530</td>
<td>2.088</td>
<td>.136</td>
</tr>
<tr>
<td>Error (Room*Condition)</td>
<td>13119.929</td>
<td>46</td>
<td>285.214</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS<em>Room</em>Condition</td>
<td>138.426</td>
<td>4</td>
<td>34.606</td>
<td>0.300</td>
<td>.877</td>
</tr>
</tbody>
</table>

with indirect comparisons, suggesting insensitivity to loudspeaker differences in this condition. It was also noted that, generally only a contraction of ratings, not a change in the order of preference between conditions, is seen. The loudspeaker*condition interaction may therefore be described as ordinal. This means that a straightforward interpretation of the loudspeaker*condition interaction can be made: it signifies a reduction in differences between loudspeakers in the indirect comparison condition compared to the direct comparison condition and it can be concluded that this is evidence for compensation affecting the loudspeaker factor. However, it should be noted that the effect of comparison method was shown to be driven by a reduction in perceived difference between LS1 and the other loudspeakers (see Figure 3.4). Therefore, the loudspeaker by condition interaction is mainly based on a contraction of ratings for only one loudspeaker. A similarly large movement for all loudspeakers would provide stronger evidence of compensation.

**Hypothesis 1c summary**

The significant ordinal loudspeaker*condition interaction is evidence of a reduction in variation between loudspeaker models with indirect comparisons. This shows compensation affects the loudspeaker differences when indirect comparisons are made. This effect appears to be mainly due to LS1 being perceived as less extreme when
indirect comparisons are made. Results would gain validity if this pattern of contraction was replicated with all 3 loudspeakers. The interaction is not quite significant at \( p < .05 \), but is significant at \( p < .10 \). Together with Olive et al’s results, these results are suggestive of compensation for the loudspeaker factor when listening to music but not conclusive. Compensation for rooms could not be tested as the rooms were not perceived to be significantly different with direct comparisons, so could not reduce further with compensation.

**Hypothesis 1d—Room*loudspeaker interactions**

Table 3.6 shows a significant room*loudspeaker interaction in the loudspeaker condition \( F(4, 46) = 4.232, \ p = .003 \) (\( p < .001 \) with unreliable participants removed) and in the room condition \( F(4, 46) = 6.139, \ p < .001 \) (\( p = .001 \) with unreliable participants removed). These room*loudspeaker interactions show that the indirect factor is still perceived and that compensation is not complete. However, the exact nature of the interaction can affect the interpretation of the compensation seen in this experiment.

**Disordinal room*loudspeaker interactions**

The ANOVA analysis does not distinguish between ordinal and disordinal interactions when calculating interaction terms. However, the distinction may be important when analysing compensation. A disordinal room*loudspeaker interaction occurs when the order of preference for loudspeakers changes between rooms. Only one instance of a disordinal interaction can be seen in Figures 3.4 and 3.6, where LS3 changes from the 2\(^{nd}\) most preferred loudspeaker in R1 and R2 to most preferred in R3 (see Figure 3.4). This occurs in both direct and indirect conditions. This result shows that the room in which the loudspeaker is heard has the effect of changing the sound of the loudspeaker in a positive/negative way to result in increased/decreased ratings for that same loudspeaker. i.e. ratings of LS3 depend on the room (or the room ratings depend on the loudspeaker). Changes in loudspeaker preference between rooms have been found in a number of previous listening tests where direct comparisons were made (Toole 2006) and are expected where direct comparisons are made. However, when comparisons are indirect and compensation occurs, little or no loudspeaker*room
interaction is expected. The presence of disordinal interactions shows compensation is not complete because the indirect factor is still having an effect on ratings. Both disordinal interactions and the main effects for the indirect factor would be expected to reduce to non-significance where compensation for that factor is complete. A reduction in disordinal interactions cannot be tested in this study because the paradigm does not allow for this. However, the size of the interaction can be examined. Disordinal room*loudspeaker interactions in this study are very small. This is an indication that the indirect factor is not having a large effect on ratings via interactions, and that compensation has occurred. A test of whether disordinal interactions decrease with increased time gaps should be considered as method of measuring compensation in future studies.

**Ordinal room*loudspeaker interactions**

Ordinal room*loudspeaker interactions occur in both conditions and appear to make up most of the contribution to the significant room*loudspeaker interaction effect. Ordinal interactions show that the degree of difference in preference between a factor (e.g. between loudspeakers) is larger at some levels of the other factor than others (e.g. larger within some rooms compared to others) but the order of preference is the same. For example in Figure 3.4 it can be seen that the variation between loudspeakers is larger in R3 than R2, and larger in R2 than in R1, but that the order of preference for loudspeakers stays the same in each room (except for the disordinal interaction described above). Like with disordinal interactions, ordinal interactions may occur because the loudspeakers and rooms combine to create more positive or more negative perceptions. In this case ordinal interactions can be interpreted in the same way as disordinal interactions and show that the indirect factor is continuing to have an influence on ratings.

However, ordinal interactions can also be interpreted as showing a different effect. Ordinal interactions might indicate that the effect of one factor on preference is larger or smaller at some levels of the other factor than others. For example the differences between loudspeakers are larger in R3 than in R1 in this study. This may occur because of differences in the amount of compensation within the levels of the factor that is being compensated for. E.g. loudspeaker difference might be larger in R3 because more
compensation for the room has occurred here leaving the listener more sensitive to the difference between loudspeakers within that room. It is not clear why some rooms may undergo more compensation than others (or some loudspeakers undergo more compensation than others). A possible reason may be that R3 has a stronger effect on perceptions to begin with, overpowering the loudspeaker differences, so compensation might be larger here. Therefore, the presence of ordinal interactions may be evidence that the indirect factor is still having an effect in some instances, but this varies, and in some instances its effects might be small, or have reduced completely. A condition where loudspeaker and room changes occur in a completely intermixed fashion might be useful in any future testing that aims to separate the two explanations of ordinal interactions. The loudspeaker*room interactions seen in this experiment appear to be largely ordinal but they are a mix of both ordinal and disordinal to some extent, and all that can be concluded with the current paradigm is that the presence of any such interactions shows non-complete compensation.

**Hypothesis 1d summary and discussion**

A significant reduction in main effect shows compensation in the presence of significant room*loudspeaker interactions but the presence of such interactions shows that compensation is not complete (or is at least is not complete at some levels of the other factor). A reduction in interactions is also expected with compensation. Because of the task set-up this could not be tested here but it might be useful to examined this reduction in future studies as an alternative means of assessing compensation. The fact that interactions were small is suggestive that compensation has occurred.

The significant interactions between loudspeakers and rooms in this study is mostly due to ordinal interactions. These interactions are likely to be a reflection of the fact that the effect of loudspeakers is larger in some rooms than others (and/or that the effect of rooms is larger for some loudspeakers). This may suggest that compensation for the indirect factor is larger at some levels of that factor than others. It may also suggest that the indirect factor is still having an effect on the direct factor, to change the sound of that factor. The reasons for loudspeaker*room interactions should be investigated further but it can be concluded that they show that compensation for the direct factor is not complete in this test.
3.5 Experiment 1: Summary and discussion

In the previous chapter it was acknowledged that further research into compensation for transmission channel colouration over the longer listening periods that occur in real-world listening is needed. An experimental paradigm by Olive et al. (1995) that appeared to offer an appropriate method for examining this was discussed. This paradigm was used in the current study to examine compensation for loudspeakers and rooms. The experiment reported here aimed to replicate the compensation for these channels seen in Olive et al.’s (1995) work. This replication is necessary to show that compensation occurs, that it can be elicited using this paradigm, and that this paradigm can be used in future tests to examine the mechanisms of real-world compensation. The current experiment used the same non-speech musical stimuli as in Olive et al.’s (1995) study but that study was extended through the use of different loudspeakers and rooms. This was done to confirm that Olive’s experiment shows a genuine compensation rather than a result that is an artefact of the stimuli used. In Experiment 1 listeners either made indirect comparisons of 3 loudspeakers and rooms, where they heard stimuli continuously for a longer period and with break periods between presentations. This condition was the same as in Olive et al.’s (1995) experiment (except 2 minute gaps rather than approximately 10 minute gaps were used and shorter continuous listening periods were used). This condition represented a real-world listening scenario. Alternatively, they made direct comparisons whereby they listened to stimuli for short periods and made side-by-side comparisons, with no breaks, as is common in laboratory listening.

Descriptive measures showed the expected reduction in perceived differences between loudspeakers with indirect comparisons. This was statistically significant at the p<.10 level. This is good evidence of compensation but fails to reach the stricter p<.05 level. Results with less reliable participants removed do meet the <.05 criteria, so some further evidence for compensation for the loudspeaker factor is provided by this study. The differences between rooms in this study were not large enough for listeners to reliably tell the difference between them even when the direct comparison method was used. Therefore, compensation could not occur to a significant level. However, descriptive statistics do show some tendency for compensation with indirect room
comparisons. Non-significant differences between rooms with direct comparisons may be because the physical range in rooms is small, factors prevented listeners hearing the difference between rooms in this experiment, or factors made listeners more unreliable in rating rooms. Alternatively, it may be that compensation for the room occurred immediately and was at work even when direct comparisons were made. It should be ensured that both rooms and loudspeakers can be easily differentiated with direct comparisons in future studies, as it is not possible to examine compensation where the variation in stimuli is too small.

The lack of large and significant compensation for the loudspeaker factor may be explained by the reduced continuous listening length and time gaps for the indirect comparison condition compared to those used in Olive et al.’s (1995) study. This failure to replicate may show the importance of a longer time course of listening to compensation. Overall this experiment, together with the results of Olive et al. shows some evidence of compensation when listening to music through loudspeakers and rooms over a time course more representative of real-world listening. However, further work is needed to confirm that this occurs to a significant level for both loudspeakers and rooms.

As compensation was not fully significant the magnitude of compensation is not assessed in detail. If no compensation occurred there would be no reduction in loudspeaker and room differences with indirect comparisons. If complete compensation occurred results would show a complete reduction of loudspeaker/rooms differences in the indirect comparison condition. A result between these extremes was seen for the loudspeaker factor (and a tendency towards this is evident for the room factor). Standard deviation ratios appear to show compensation that is small to moderate. Though some compensation was shown for the loudspeaker factor it is clear that the difference due to the loudspeaker was still perceived when compared indirectly, so this compensation cannot be considered complete over the time course tested. Compensation for both rooms and loudspeakers may be larger in future experiments when longer time courses and/or a wider range of stimuli are used.

The time course of effects was not specifically studied in this experiment but it may be that compensation is related to the longer time course of listening with
indirect comparisons. This may mean that, as Olive et al. suggest, the time gap between stimuli is important due to memory or other time-gap sensitive mechanisms causing compensation. The result could also mean that longer listening brings about compensation. The fact that reduced compensation was seen here compared to in Olive et al.’s study may be because the longer listening periods in Olive et al.’s study are necessary to bring about significant compensation. Evidence of compensation in this musical context along with that in Olive et al.’s study shows that the processes that are at work in this study are of a general auditory nature and may work to compensate for the spectrum of any channel when listening to any sound.

The measures of compensation (loudspeaker*condition interactions and room by condition interactions) do not consider whether or not the change in ratings is ordinal resulting in a contraction/expansion of ratings (as is expected with compensation) or, whether they are disordinal which is not expected with compensation and indicates a lack of compensation. The pattern of movement of data between conditions must be examined by eye to confirm this. The graphs in this study show that the main effect by condition interactions were mainly ordinal and so a straightforward interpretation of effects could be made. But the inclusion of any disordinal interaction in main effect*condition measure means that this measure may not simply be measuring compensation. A numerical analysis that accounts for the direction of change might be preferable in future experiments, if increased disordinal interactions are present.

It might also be of interest to examine reductions in room*loudspeaker interactions in future research to offer an alternative measure of compensation. A reduction in these interactions would be expected if compensation for the indirectly compared factor occurred. The size of interactions before and after compensation could not be compared in the study because of the format of the paradigm. However, generally, the size of the room*loudspeaker interactions after compensation can be examined. These interactions were significant showing that compensation was not complete, but they were small suggesting some compensation occurred. This result contrasts with many laboratory studies that test listening over short time courses and show large interactions. Large room*loudspeaker interactions seen in laboratory studies indicate that results may not be representative of the compensation that occurs when listening over longer real-world time courses.
Where room*loudspeaker interactions are disordinal it is likely that they show that the indirect factor is still causing a change in preference ratings. A similar interpretation of ordinal interactions might be taken. However, another explanation of ordinal interactions is possible; ordinal interactions might show that compensation for the indirect factor is uneven (e.g. there is more compensation at some levels of the indirectly compared factor than others). This interpretation can be tested by using a fully intermixed condition in future work. Either way, no interactions would be expected where compensation is complete, so any room*loudspeaker interaction shows that an effect of the indirect factor remains. Future studies examining room and loudspeaker compensation should ideally be set up so that a change in interactions with compensation can be measured. However, no specific testing of reductions in loudspeaker*room interactions will be conducted as part of this thesis, as it is not possible to measure this using the paradigm selected. Only comments on the size of these interactions will be made.

A number of improvements to the study can be made to increase the likelihood of finding significant effects when stimuli are compared directly and significant compensation in the future. A wider range of stimuli should be used to prevent floor effects in measuring compensation. However, it may not be desirable to increase the range greatly as even though the range in rooms might be small, it may be said that the rooms and loudspeakers chosen for this study represent the range commonly found in the real world. Different programme material may also be useful in increasing the perceived differences between the particular loudspeakers and rooms used. Longer time gaps, like those used in Olive et al.’s study, and increased continuous listening to the indirect stimuli should also be used to elicit increased compensation. Increased power to detect small differences by increasing the number of participants and taking more repeated ratings will also help to show statistically significant compensation where the effect is small. Efforts should also be made to decrease measurement error. This may be achieved via familiarisation phases. Alternatively, analysis on a participant by participant basis might increase the magnitude of the compensation effect found here. Some sensitivity for detecting an effect of comparison method was lost when ratings of loudspeaker preference and room preference were measured across participants, as each participant may prefer a different loudspeaker/room. However, it is usual to average
preference ratings for loudspeakers and rooms across participants, as was done in this test, to get more valid measures of loudspeaker/room preference. The results with such analysis are more conservative but an alternative method could be considered in future if compensation effects continue to appear to be present but small.

3.6 Experiment 1: Conclusion

The experiment presented in this section aimed to replicate the experiment by Olive et al. (1995) and confirm compensation for spectral colouration caused by loudspeakers and rooms. This experiment tested compensation with the same musical stimuli used in Olive et al.’s original experiment but used different rooms and loudspeakers in an attempt to extend the validity of that research. A confirmation of compensation with Olive et al.’s (1995) paradigm is necessary if this paradigm is to be used to determine the mechanisms behind compensation in future tests.

The current experiment did reveal compensation for the loudspeaker factor but this was only significant at the $p<.10$ level. There was an indication of a trend toward compensation for the room factor but stimuli differences were not large enough to be statistically significant when stimuli were compared directly over a short listening time course, so a further significant reduction with a longer listening time course could not be measured. This floor effect prevents a firm conclusion regarding compensation for the room factor. Overall it may be concluded that this experiment shows some compensation and this result indicates that compensation is unlikely to be an artefact of Olive et al.’s precise method or stimuli. However, until significant compensation for both factors at $p<.05$ is obtained it cannot be confirmed that real-world compensation occurs using this paradigm.

The fact that only small, non-significant, differences between rooms were observed when direct comparisons were made appears to be partly due to a small physical range of rooms and/or other factors of the experiment that meant that it was difficult to hear differences between rooms. A wider range of rooms, and/or conditions more conducive to hearing room differences, should be used in future work. Another reason for the reduced compensation might be the smaller difference in the listening time
course between direct and indirect conditions in this study compared to Olive et al.’s study. The break phase in the indirect comparison condition was reduced from that used in Olive’s study and the length of listening was also reduced. Increasing the time course of listening in the indirect comparison condition back to the duration used in Olive et al.’s work may result in larger compensation if time course is important to compensation.

The presence of significant room*loudspeaker interactions show that compensation was not complete. Some room*loudspeaker interaction is expected as only significant compensation, rather than complete compensation, is expected in this test. Disordinal room*loudspeaker interactions can be interpreted simply and show that the indirect factor is still affecting ratings but ordinal interactions may show that compensation for the indirect factor is occurring but to a different extent depending on the specific channel that is being compensated for (e.g. compensation may be greater for room 1 compared to room 2). A reduction in either type of room*loudspeaker interaction with longer listening would show that compensation has occurred but it was not possible to measure this reduction using this paradigm. Future tests should consider a way in which a reduction in room*loudspeaker interactions can be used to measure compensation. However, the tests in the remainder of this thesis will continue to primarily examine compensation using loudspeaker by condition and room by condition interactions. It should be acknowledged that these interactions might also be disordinal and disordinal interactions would not indicate compensation. Therefore, the interpretation of compensation using this paradigm is more complex than suggested by Olive et al.’s (1995) original study. If large disordinal factor by condition interactions are seen in future tests a method of analysis that does not include these should be considered. Like Olive et al.’s original study the current study did not aim to determine the mechanisms of real-world compensation.

It is concluded that together with Olive et al.’s original research the almost significant trend towards compensation shows that compensation for loudspeaker factor occurs in a real-world scenario. The results for the room factor are not conclusive as floor effects prevented compensation from being measured. Compensation for loudspeakers and rooms should be tested further with conditions more conductive to compensation to confirm that this occurs for either or both channels. Further experiments will be
conducted to confirm compensation with this paradigm.

3.7 Experiment 2: Research aims and hypotheses

The previous experiment replicated the experiment by Olive et al. (1995) by matching the materials, methods and analysis. It also extended the research by using different loudspeakers and rooms. As was expected by the compensation hypothesis (see Section 3.1.1), the results of the replication experiment showed a tendency for listeners to report smaller variation in loudspeakers and rooms with indirect comparisons, suggesting loudspeaker and room compensation. However, effects were small and not statistically significant at the $p<.05$ level, so further confirmation of compensation is necessary.

A number of issues with the replication experiment were discussed in the previous chapter. The loudspeakers/rooms were chosen to be similar to those used in Olive et al.’s (1995) experiment and representative of consumer loudspeakers and domestic rooms. However, results showed that the rooms chosen in the replication experiment were not different enough for listeners to reliably perceive the differences between them, even when direct comparisons were made. This resulted in floor effects whereby the magnitude of perceived difference between rooms could not decrease further with compensation. Increasing the perceived room and loudspeaker range used is likely to increase the magnitude of the effect of these channels when making direct comparisons and therefore provide an increased possibility of a significant reduction in effect with indirect comparisons (i.e. floor effects will be reduced). Additionally, increasing the length of continuous listening and the length of time gaps between stimuli may result in more compensation.

The aim of the current study is to implement the changes recommended in the previous section and measure compensation when the conditions are more favourable to creating compensation effects. As in Experiment 1, the main aim of this study is to determine whether compensation occurs for both loudspeakers and rooms. Therefore, as for Experiment 1, this experiment aims to answer Research Question 1a, and the chapter questions 3.1 and 3.2.
3.7 EXPERIMENT 2: RESEARCH AIMS AND HYPOTHESES

Question 3.1: Can compensation for loudspeakers and rooms seen in the real-world experiments described in chapter 1 be elicited again with new stimuli?

Question 3.2: Is this compensation an artefact of the experimental process or is it likely to be a genuine listening phenomenon?

The main aim of this experiment is not to determine mechanisms of compensation but a preliminary analysis of the role of listening length in compensation is conducted. This contributes to answering Research Question 1b and Chapter Question 3.3.

Question 3.3: What type of perceptual mechanisms might be behind this compensation?

3.7.1 Rationale

The main purpose of the current experiment was to increase perceived differences between loudspeakers/rooms in order to prevent floor effects and to increase the compensation. The use of a wider range of loudspeakers/rooms, in terms of objective measurements, was considered but it was decided that it would be preferable to continue to use loudspeakers and rooms similar in range to those in Olive et al.’s (1995) experiment. This is because the rooms and loudspeakers selected are representative of real-world variation in these items. Instead the perceived differences between them can be maximised with other methods such as the programme item that is played through them.

Using a speech programme item: Olive et al. (1995) examined only musical programme items. However, research suggests that compensation may be more likely to occur with speech stimuli (Fowler 2006, Liberman et al. 1967, Repp 1982, Watkins 1991). Anechoic male speech was used in this experiment. This programme item was also chosen because a pilot test showed moderate to large differences between the same loudspeakers and rooms used in Experiment 1 being perceived, therefore reducing the likelihood of floor effects affecting the current experiment.

Increasing time gaps: Olive et al. (1995) reported time gaps of approximately 10 minutes between presentations of indirectly compared stimuli. The previous replication experiment reduced time gaps to 2 minutes to reduce the overall length of the
3.7 EXPERIMENT 2: RESEARCH AIMS AND HYPOTHESES

Experiment. If compensation is due to time gaps these gaps may not have been sufficient to observe an effect. The current experiment increased time gaps to 10 minutes. This should have the effect of increasing the magnitude of any compensation that is to do with time gaps between 2 and 10 minutes.

**Increasing time listening to the indirect channel:** Olive et al.’s (1995) experiment involved listeners conducting multiple ratings for the directly compared factor whilst listening to the indirect factor continuously for periods up to a few minutes (exact time not reported). In Experiment 1 listeners only heard the indirect factor continuously for 25 seconds (average time). If an effect of comparison method is due to continuous listening to the indirect factor, then it may be necessary to extend listening periods. It is estimated that time spent listening to the indirect factor was nine times longer in Olive et al.’s (1995) work (in Olive et al. each rating was repeated for 3 programme items and 3 loudspeaker positions). The time spent listening to the indirect factor was lengthened by a factor of 9 in the current experiment by introducing more repeated ratings at each level of the indirect factor (e.g. loudspeakers were rated more times within each room).

The above methods aimed to increase the magnitude of perceived differences between loudspeakers/rooms and the magnitude of any compensation i.e. they increase effect sizes within the experiment. Additional improvements to the study were made to increase the likelihood of finding significant effects by reducing error in measurements. These were: increasing the number of participants to increase statistical power, increasing the number of repeated ratings to increase statistical power, and using a longer familiarisation phase to improve the reliability of loudspeaker and room ratings. A familiarisation phase appears not have been present in Olive et al.’s (1995) work but this phase is likely to improve reliability of ratings. This phase may cause compensation to occur before the test phases have begun, but this is unlikely as long as there is a suitable break before the beginning of the test.
3.7.2 Hypotheses

The main hypothesis to be tested in this experiment is the same as in Experiment 1. This is Hypothesis 1 set out in Section 3.1.1, namely compensation with indirect comparisons, as seen in Olive et al.’s (1995) experiment, will occur. This time only a selection of the sub-hypotheses from the previous experiment (see Section 3.1.1) were tested. It was established that the absolute variation in loudspeakers and rooms was not expected to be similar to those in Olive’s experiment so Hypotheses 1a and 1b were not tested. However, an additional hypothesis similar to Hypothesis 1a was tested. This is labelled Hypothesis 1a(ii). The Hypotheses 1c and 1d from the previous experiment were tested. An additional hypothesis, Hypothesis 1f, was tested. These sub hypotheses are described below.

Hypothesis 1(aii)

A significant effect for both the loudspeaker and room factors will occur when direct comparisons are made so that significant compensation with indirect comparisons can occur.

Hypothesis 1(c)

As in Experiment 1, significant ordinal loudspeaker*condition and significant ordinal room*condition interactions representing a significant decrease in variation when a factor is compared indirectly are expected.

Hypothesis 1(d)

Room*loudspeaker interactions are not expected to occur in either condition where compensation is complete. The presence of significant room*loudspeaker interactions shows that compensation is not complete (at least not for all rooms/loudspeakers). Room*loudspeaker interactions are expected to be small or non-existent.

Hypothesis 1(e)

This hypothesis was not tested in the previous experiment. It is expected that increased compensation will be observed when the participant has spent longer listening to
the indirect factor continuously. The effect of comparison method is expected to be greater when the final 3 repetitions within each block (experiment pages 7-9) are examined compared to when the first 3 repetitions (pages 1-3) are examined. If compensation occurs with increased listening to the indirect factor, it is also expected that room*loudspeaker interactions (tested under Hypothesis 1d) will be larger in analysis consisting only of the first 3 experimental pages compared to the final three. A difference in interactions between the two groups is examined.

Other than the changes to the stimuli and listening lengths described above the experiment aims to replicate the procedure used in Olive et al.’s (1995) work. The task procedure and any variation in the task from that described in Experiment 1 (Section 3.3), is described in Section 3.9. The results are presented in Section 3.10. The extent of compensation for loudspeakers and rooms and possible mechanisms behind this are discussed further in the discussion and conclusion section (Section 3.11).

3.8 Experiment 2: Participants and material

Participants were selected using the same selection criteria as in the previous experiment (see Section 3.2.1). The loudspeakers and rooms used were also the same as in the previous experiment (Section 3.2.3). The method of stimulus production is reported in Section 3.2.3. In this experiment each loudspeaker/room impulse response was convolved with the new male speech programme item using Matlab. This item was anechoic male speech taken from a sample from the Bang and Olufsen Archimedes research project stimulus CD:


The programme item was looped and cross-faded for continuous playback (see Section 3.2.3).
3.9 Experiment 2: Instructions and procedure

Participants were seated at a laptop computer with Sennheiser HD 600 headphones and allowed time to read instructions. Instructions were identical to those given in the previous experiment (see Section 3.2.5 and Appendix A). Participants were told that the experiment would consist of 2 sessions, ‘condition 1’ and ‘condition 2’ (i.e. the loudspeaker and room conditions) and that condition 1 and condition 2 would be separated by 24 hours or longer. There were four phases within each condition (1 familiarisation and 3 test phases) each separated by 10 minute break periods. Participants completed one condition of the experiment (either the room or the loudspeaker condition, randomly assigned and single blind). After at least 24 hours they returned to complete the 2nd condition. As for Experiment 1 all listening was done using binaural recordings of loudspeakers being played in rooms rather than in-situ listening. Listening was blind.

Familiarisation phase

Listeners began each condition with a familiarisation phase. The structure of this phase for participants in the loudspeaker condition and in the room conditions can be seen in Table 3.9. Listeners had to audition and make ratings for stimuli presented in the same manner as in the test phases but here the full range of stimuli were presented in close succession to give the listener experience with the range of stimuli in the test. This phase was three times longer than in the previous experiment to increase practice time (each stimulus was heard three times rather than only once). After listeners completed the familiarisation phase there was a 10 minute break before the first test phase.

Test phases

The test phase for each condition was separated into three blocks with 10 minute break periods between each block (see Table 3.10).
Table 3.9: Familiarisation Phase. Stimulus presentation order is shown for a single participant. The structure of the phase for the loudspeaker and the room condition is presented. Loudspeaker Assignment Order describes the random assignment of the levels of the directly compared factor to the buttons A, B and C on the computer user interface.

<table>
<thead>
<tr>
<th>Experimental Page</th>
<th>Block</th>
<th>Room</th>
<th>Loudspeaker Assignment Order (A, B, C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>R2</td>
<td>LS1, LS2, LS3</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>R1</td>
<td>LS1, LS3, LS2</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>R3</td>
<td>LS3, LS2, LS1</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>R3</td>
<td>LS1, LS2, LS3</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>R1</td>
<td>LS1, LS2, LS3</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>R2</td>
<td>LS2, LS3, LS1</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>R1</td>
<td>LS2, LS1, LS3</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>R2</td>
<td>LS2, LS1, LS3</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>R3</td>
<td>LS1, LS2, LS3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experimental Page</th>
<th>Block</th>
<th>Loudspeaker</th>
<th>Room Assignment Order (A, B, C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>LS1</td>
<td>R2, R3, R1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>LS2</td>
<td>R1, R2, R3</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>LS3</td>
<td>R3, R1, R2</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>LS2</td>
<td>R2, R1, R3</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>LS3</td>
<td>R3, R1, R2</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>LS1</td>
<td>R1, R2, R3</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>LS3</td>
<td>R1, R3, R2</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>LS1</td>
<td>R1, R2, R3</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>LS2</td>
<td>R2, R1, R3</td>
</tr>
</tbody>
</table>
Table 3.10: The procedure (both conditions) for a single participant starting in the loudspeaker condition. This listener completes the loudspeaker condition before completing the room condition after at least 24 hours. Loudspeaker and room assignment order describes the random assignment of the levels of the directly compared factor to the buttons A, B and C on the computer user interface.

<table>
<thead>
<tr>
<th>Loudspeaker Condition</th>
<th>Room Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Page</td>
<td>Block</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
</tr>
</tbody>
</table>

10 minute break

<table>
<thead>
<tr>
<th>Page</th>
<th>Block</th>
<th>RM</th>
<th>LS Assignment (A, B, C)</th>
<th>24h</th>
<th>Page</th>
<th>Block</th>
<th>LS</th>
<th>RM</th>
<th>Assignment (A, B, C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>R3</td>
<td>LS1, LS3, LS2</td>
<td></td>
<td>1</td>
<td>2</td>
<td>LS3</td>
<td>R2</td>
<td>R3, R1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>R3</td>
<td>LS1, LS2, LS3</td>
<td></td>
<td>2</td>
<td>2</td>
<td>LS3</td>
<td>R1</td>
<td>R2, R3</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>R3</td>
<td>LS3, LS2, LS1</td>
<td></td>
<td>3</td>
<td>2</td>
<td>LS3</td>
<td>R3</td>
<td>R1, R2</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>R3</td>
<td>LS2, LS1, LS3</td>
<td></td>
<td>4</td>
<td>2</td>
<td>LS3</td>
<td>R2</td>
<td>R1, R3</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>R3</td>
<td>LS1, LS2, LS3</td>
<td></td>
<td>5</td>
<td>2</td>
<td>LS3</td>
<td>R3</td>
<td>R1, R2</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>R3</td>
<td>LS2, LS3, LS1</td>
<td></td>
<td>6</td>
<td>2</td>
<td>LS3</td>
<td>R1</td>
<td>R2, R3</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>R3</td>
<td>LS1, LS2, LS3</td>
<td></td>
<td>7</td>
<td>2</td>
<td>LS3</td>
<td>R2</td>
<td>R3, R1</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>R3</td>
<td>LS2, LS1, LS3</td>
<td></td>
<td>8</td>
<td>2</td>
<td>LS3</td>
<td>R1</td>
<td>R2, R3</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>R3</td>
<td>LS1, LS2, LS3</td>
<td></td>
<td>9</td>
<td>2</td>
<td>LS3</td>
<td>R1</td>
<td>R3, R2</td>
</tr>
</tbody>
</table>

10 minute break

<table>
<thead>
<tr>
<th>Page</th>
<th>Block</th>
<th>RM</th>
<th>LS Assignment (A, B, C)</th>
<th>24h</th>
<th>Page</th>
<th>Block</th>
<th>LS</th>
<th>RM</th>
<th>Assignment (A, B, C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>R1</td>
<td>LS2, LS1, LS3</td>
<td></td>
<td>1</td>
<td>3</td>
<td>LS2</td>
<td>R2</td>
<td>R3, R1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>R1</td>
<td>LS1, LS3, LS2</td>
<td></td>
<td>2</td>
<td>3</td>
<td>LS2</td>
<td>R1</td>
<td>R2, R3</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>R1</td>
<td>LS3, LS2, LS1</td>
<td></td>
<td>3</td>
<td>3</td>
<td>LS2</td>
<td>R2</td>
<td>R3, R1</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>R1</td>
<td>LS3, LS2, LS1</td>
<td></td>
<td>4</td>
<td>3</td>
<td>LS2</td>
<td>R2</td>
<td>R1, R3</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>R1</td>
<td>LS1, LS2, LS3</td>
<td></td>
<td>5</td>
<td>3</td>
<td>LS2</td>
<td>R2</td>
<td>R3, R1</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>R1</td>
<td>LS2, LS3, LS1</td>
<td></td>
<td>6</td>
<td>3</td>
<td>LS2</td>
<td>R3</td>
<td>R1, R2</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>R1</td>
<td>LS2, LS1, LS3</td>
<td></td>
<td>7</td>
<td>3</td>
<td>LS2</td>
<td>R2</td>
<td>R3, R1</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>R1</td>
<td>LS1, LS2, LS3</td>
<td></td>
<td>8</td>
<td>3</td>
<td>LS2</td>
<td>R1</td>
<td>R2, R3</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>R1</td>
<td>LS2, LS1, LS3</td>
<td></td>
<td>9</td>
<td>3</td>
<td>LS2</td>
<td>R1</td>
<td>R2, R3</td>
</tr>
</tbody>
</table>
At the beginning of the experiment listeners were randomly assigned to completing the loudspeaker or room condition first. A description of the test phase for a participant starting in the loudspeaker condition is given:

**Block 1, Experimental page 1:** The listener was presented with the first experimental page, which was identical in appearance to the interface used in the previous experiment (Appendix B). A room was selected from the three possible rooms (e.g. R2) using a Latin squares method. Recordings of the speech stimulus being played through each of the loudspeakers (LS1, LS2 and LS3) in that room were randomly assigned to buttons A, B and C. For the reasons discussed in Section 3.2.5, listeners were not informed about what sort of stimuli they were listening to. Listeners were instructed to rate each stimulus for preference. Listeners could switch immediately between stimuli by pressing buttons A, B and C in any order to make comparisons. They could listen to each stimulus as many times as they desired before making a rating. As in Experiment 1, listeners listened to the loudspeakers for milliseconds or seconds before switching to another to make a comparisons. Once ratings were made they moved on to experimental page 2.

**Block 1, Experimental pages 2-9:** Page 2 presented the listener with the same stimuli (speech through LS1, LS2 or LS3) heard in the same room as previously. However, the stimuli were newly randomly assigned to buttons A, B and C to attempt to prevent order effects on ratings and listeners noticing that they were preforming repeated ratings of the same stimuli. The listeners made ratings for each loudspeaker and moved on to page three. This process continued until 9 experimental pages were completed. This provided 9 replicate ratings of each loudspeaker within the same room. During the block the listener was exposed to the indirect factor continuously for nine experimental pages (this contrasts with the previous experiment where the room was changed every experimental page).

**Break Period:** After the first block there was a 10 minute break period. This follows the same procedure as the 2 minute breaks in the previous experiment (see Section 3.3).

**Block 2:** In the next test block the procedure was identical the first block except this time the listener heard the loudspeakers presented in a different room (e.g R3)
throughout the block.

**Break period**: 10 minute break period.

**Block 3**: In the 3\textsuperscript{rd} test block the loudspeakers were rated in the final room (e.g. R1).

**Room condition**: The participant returned at least 24 hours later to complete the room condition. An identical procedure was followed for the room condition except that this time rooms were directly compared and loudspeakers were indirectly compared. In the first block a loudspeaker was selected on a Latin squares basis (e.g LS1). A stimulus consisting of the speech item played through that loudspeaker in each room (R1, R2 or R3) was randomly assigned to buttons A, B, C. The listener rated each room and repeated this for 9 pages. The loudspeaker remained the same within a block. In blocks 2 and 3 the remaining two loudspeaker were presented.

The time spent in each block (i.e. the time spent listening to the indirect factor), averaged across participants is given in Table 3.11 for the loudspeaker condition and room condition.

<table>
<thead>
<tr>
<th>Loudspeaker</th>
<th>Room</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block</td>
<td>Mean (mins)</td>
</tr>
<tr>
<td>Familiarisation</td>
<td>7 mins</td>
</tr>
<tr>
<td>1</td>
<td>6 mins</td>
</tr>
<tr>
<td>2</td>
<td>6 mins</td>
</tr>
<tr>
<td>3</td>
<td>4 mins</td>
</tr>
</tbody>
</table>

### 3.10 Experiment 2: Results

This section reports statistical analysis to check reliability and ANOVA assumptions (Section 3.10.1), descriptive measures (Section 3.10.2) and statistical analysis to examine compensation (Section 3.10.3).
3.10 EXPERIMENT 2: RESULTS

3.10.1 Checking for reliability and assumptions of repeated measures ANOVA

Data was checked for normality using Q-Q plots and the reliability of participants assessed. Residuals were normally distributed around the mean for each loudspeaker/room combination, for each participant. No data was removed because of unreliability. Where Mauchly’s test of sphericity indicated that the data was non-spherical Huynh-Feldt adjusted degrees of freedom were used to produce variance measures.

3.10.2 Descriptive measures

This section examines graphs and descriptive statistics with the aim of assessing whether the variation in loudspeakers and rooms is larger when direct comparisons are made compared to indirect comparisons.

Loudspeaker factor

Figure 3.8 shows the mean rating for each loudspeaker in each room (averaged across all participants and repeats). The results show an increase in variation in loudspeaker ratings in the loudspeaker condition (direct comparisons) compared to the room condition (indirect comparisons).

The difference between loudspeakers is larger in the direct comparison condition for all rooms. However, the difference between loudspeakers is driven by LS3 being more preferred than the other 2 loudspeakers in all rooms. There may not be any significant difference between LS1 and LS2 except in R2.

Room*loudspeaker interactions are small. There is a small ordinal movement where loudspeakers differences are slightly larger in R2 compared to R1 and R3 for both conditions. The order of loudspeaker preference does not appear to change significantly between conditions confirming stability of preference (averaged across participations) between conditions.
3.10 EXPERIMENT 2: RESULTS

Figure 3.8: Mean preference score for each loudspeaker, within each room, in the Loudspeaker condition (A). The mean score for each loudspeaker, within each room, in the room condition (B). Differences in loudspeakers are shown as differences between lines.

Variance measures

The variance of mean loudspeaker ratings within each room is also shown in Table 3.12. The variation in loudspeakers is largest in R2 and smallest in R1 in both conditions. Variation in loudspeakers is larger when direct comparisons are made compared to indirect comparisons for all rooms. Standard deviation ratios show that there is moderately larger loudspeaker variation in the direct comparison condition for all rooms and that the size of compensation is slightly different for each room. The effect of condition appears to be smallest for R1 ($SD_d/SD_i = 4.68/1.38 = 2.54$) and largest for R2 ($11.24/4.42 = 3.39$).

Main Effects

Figure 3.9 shows the loudspeaker main effect in both conditions. An ordinal loudspeaker by condition interaction can be seen. The interaction moves in the direction expected by the compensation hypothesis: the effect of the loudspeaker factor reduces with indirect comparisons (i.e. in the room condition). Interaction effects are tested for statistical significance in Section 3.10.3. It can also been seen in Figure 3.9(A) that loudspeaker variation is mostly influenced by LS3, the most preferred loudspeaker. Its score is 62.03 points on the preference rating scale. This is 13 points greater than the least preferred loudspeaker (LS1), showing a moderate difference between loudspeakers,
Table 3.12: Loudspeaker means within each room, loudspeaker variance within each room.

<table>
<thead>
<tr>
<th>Loudspeaker</th>
<th>Condition</th>
<th>Room</th>
<th>Condition Mean</th>
<th>Room Condition Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>LS1 Mean</td>
<td>LS (Direct)</td>
<td>R1</td>
<td>47.66</td>
<td>53.17</td>
</tr>
<tr>
<td>LS2 Mean</td>
<td></td>
<td>R2</td>
<td>46.29</td>
<td>52.74</td>
</tr>
<tr>
<td>LS3 Mean</td>
<td></td>
<td>R3</td>
<td>55.00</td>
<td>64.27</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td><strong>49.65</strong></td>
<td><strong>56.73</strong></td>
</tr>
<tr>
<td>Variance</td>
<td></td>
<td></td>
<td><strong>21.93</strong></td>
<td><strong>63.67 (SD = 7.98)</strong></td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td></td>
<td>4.68</td>
<td>6.54</td>
</tr>
</tbody>
</table>

| LS1 Mean    | Room (Indirect) | R1    | 31.35          | 65.90                   |
| LS2 Mean    |               | R2    | 33.78          | 69.71                   |
| LS3 Mean    |               | R3    | 31.43          | 66.99                   |
| Mean        |               |       | **32.19**      | **67.53**               |
| Variance    |               |       | **1.91**       | **3.85**                |
| SD          |               |       | 1.38           | 1.96                    |

SD Ratio: 2.54 3.39 3.33

according to the rating scale intervals (see section 3.2.2). The difference between LS1 and LS2 is negligible (1 point) in the loudspeaker condition and differences between all loudspeakers are negligible in the room condition (3 points between the most and least preferred loudspeakers), as can be seen in Figure 3.9(B).

![Figure 3.9: Main effects for the loudspeaker factor in the Loudspeaker condition; loudspeaker means averaged across all rooms (A) and the Room condition (B).](image-url)
Room factor

Figure 3.10 shows the mean ratings for each room for each loudspeaker (averaged across participants and repeats). The results show an increase in variation between rooms in the room condition compared to the loudspeaker condition.

![Figure 3.10: Mean preference scores for each room, for each loudspeaker, in the Loudspeaker condition (A) and for each room, for each loudspeaker, in the Room condition (B). Differences in rooms are shown as differences between lines.](image)

Variance measures

The variation in mean room rating for each loudspeaker, in each condition, is shown in Table 3.13. The variation in rooms is largest for LS3 in both conditions. Standard deviation ratios show that there is larger variation in rooms in the direct comparison condition for all loudspeakers. The size of the effect of comparison method is different for each loudspeaker. The effect of condition appears to be largest for LS2 ($SD_d/SD_i = 18.66/3.28 = 5.69$) and smallest for LS3 ($19.66/6.22 = 3.16$).

Main Effects

Figure 3.11 shows the room main effect in both conditions. There is a clear room*condition interaction in the direction expected by the compensation hypothesis. The ratings for each room are reduced in the indirect comparison condition. The movement is ordinal with rooms remaining in the same position relative to each other in both conditions. The effect of room in both conditions is mainly influenced by R1, the least preferred room but there are differences between the other rooms. The difference
3.10 EXPERIMENT 2: RESULTS

Table 3.13: Room means for each loudspeaker and room variance for each loudspeaker.

<table>
<thead>
<tr>
<th>Room</th>
<th>Condition</th>
<th>Loudspeaker</th>
<th>Condition mean</th>
<th>Condition Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1 mean</td>
<td>Room (Direct)</td>
<td>LS1 LS2 LS3</td>
<td>31.35 33.78 31.43</td>
<td>311.92 (SD = 18.98)</td>
</tr>
<tr>
<td>R2 mean</td>
<td>Room (Direct)</td>
<td>LS1 LS2 LS3</td>
<td>55.02 60.53 63.76</td>
<td>348.99 (SD = 18.98)</td>
</tr>
<tr>
<td>R3 mean</td>
<td>Room (Direct)</td>
<td>LS1 LS2 LS3</td>
<td>65.90 69.71 66.99</td>
<td>386.69 (SD = 18.98)</td>
</tr>
<tr>
<td>R1 mean</td>
<td>Room (Indirect)</td>
<td>LS1 LS2 LS3</td>
<td>50.76 54.68 54.06</td>
<td>348.99 (SD = 18.98)</td>
</tr>
<tr>
<td>R2 mean</td>
<td>Room (Indirect)</td>
<td>LS1 LS2 LS3</td>
<td>55.25 50.55 66.82</td>
<td>348.99 (SD = 18.98)</td>
</tr>
<tr>
<td>R3 mean</td>
<td>Room (Indirect)</td>
<td>LS1 LS2 LS3</td>
<td>53.17 52.74 54.27</td>
<td>348.99 (SD = 18.98)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>47.66 46.29 55.00</td>
<td>48.69 49.86 62.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>45.25 50.55 66.82</td>
<td>16.47 10.76 38.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>53.17 52.74 54.27</td>
<td>4.06 3.28 6.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4.35 5.69 3.16</td>
<td></td>
</tr>
</tbody>
</table>

between the least preferred room (R1) and the most preferred (R3) is 36 rating scale points in the room condition, indicating a strong difference in preference according to the rating scale intervals. This difference is only 7 points in the loudspeaker condition, indicating a small difference in preference.

Figure 3.11: Main effects for the room factor in the Loudspeaker condition; room means averaged across all loudspeakers (A) and in the Room condition (B).

Examination of Figure 3.12 shows the loudspeaker and room main effects on the same
Figure 3.12: The main effect of for the loudspeaker factor (A) and room factor (B).

graph and movement indicative of compensation for both factors. The interaction with condition is ordinal in both cases. There is a reduction in variation of loudspeaker means in the room condition, and a reduction in variation of room means in the loudspeaker condition. The loudspeaker*condition interaction involves a small movement in preference (the difference between the most and least preferred loudspeaker increased by 10 preference points in the loudspeaker condition compared to the room condition), and the room*condition interaction involves a large movement in preference (29 preference points).

3.10.3 Statistical analysis

This section describes statistical analysis conducted to assess whether the effects described in the previous section are statistically significant and to test the specific hypotheses set out in Section 3.7.2.

Hypotheses 1a and 1c

Separate $3 \times 3$ repeated measures ANOVA analyses for each condition were conducted, see Tables 3.14-3.16. Table 3.14 shows the analysis with all the data included. Table 3.15 shows analysis conducted with only the first 3 repeats within each block of the test phase (experimental pages 1-3) and Table 3.16 shows analysis with only the final 3 repeats within each block (experiment pages 7-9). There is a significant effect of loudspeakers and rooms in both conditions both when compared directly (as
is expected by Hypothesis 1a) and when compared indirectly. There is a reduction in main effect when indirect comparisons are made. This is the case when all data is analysed and when only the first or final 3 measures from each block are analysed.

**Main effect * condition interactions**

A single ANOVA $2 \times 3 \times 3$ was conducted to examine loudspeaker*condition and room by condition interactions. See Table 3.17.
Table 3.14: Separate repeated measures of ANOVA analysis by condition. Data from experimental pages 1-9.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Type iii sum of squares</th>
<th>Df</th>
<th>Mean square</th>
<th>F</th>
<th>sig</th>
<th>Partial eta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loudspeaker</td>
<td>47138.671</td>
<td>2 (HF 1.843)</td>
<td>25580.476</td>
<td>91.195</td>
<td>.000</td>
<td>.389</td>
</tr>
<tr>
<td>Error</td>
<td>73916.884</td>
<td>286 (HF 263.515)</td>
<td>280.504</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Room</td>
<td>11114.347</td>
<td>2 (HF 1.729)</td>
<td>6428.779</td>
<td>18.008</td>
<td>.000</td>
<td>.112</td>
</tr>
<tr>
<td>Error</td>
<td>88257.875</td>
<td>286 (HF 247.224)</td>
<td>356.995</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS*Room</td>
<td>7870.009</td>
<td>4 (HF 3.645)</td>
<td>2158.853</td>
<td>13.396</td>
<td>.000</td>
<td>.086</td>
</tr>
<tr>
<td>Error</td>
<td>84008.435</td>
<td>572 (HF 521.301)</td>
<td>161.152</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Room Condition</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loudspeaker</td>
<td>3837.449</td>
<td>2 (HF 1.886)</td>
<td>2034.550</td>
<td>20.663</td>
<td>.000</td>
<td>.126</td>
</tr>
<tr>
<td>Error</td>
<td>26556.995</td>
<td>286 (HF 269.718)</td>
<td>98.462</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Room</td>
<td>298087.699</td>
<td>2 (HF 1.598)</td>
<td>186540.924</td>
<td>405.712</td>
<td>.000</td>
<td>.739</td>
</tr>
<tr>
<td>Error</td>
<td>105066.079</td>
<td>286 (HF 228.510)</td>
<td>459.787</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS*Room</td>
<td>3442.838</td>
<td>4 (HF 3.802)</td>
<td>905.646</td>
<td>8.900</td>
<td>.000</td>
<td>.059</td>
</tr>
<tr>
<td>Error</td>
<td>55315.384</td>
<td>572 (HF 543.619)</td>
<td>101.754</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3.15: Separate repeated measures of ANOVA analysis by condition. Data from experimental pages 1-3.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Type iii sum of squares</th>
<th>Df</th>
<th>Mean square</th>
<th>F</th>
<th>sig</th>
<th>Partial eta</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Loudspeaker Condition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loudspeaker</td>
<td>12260.838</td>
<td>2 (HF1.956)</td>
<td>6268.860</td>
<td>22.931</td>
<td>.000</td>
<td>.328</td>
</tr>
<tr>
<td>Error</td>
<td>25130.273</td>
<td>94 (HF91.924)</td>
<td>273.381</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Room</td>
<td>5163.366</td>
<td>2 (HF1.651)</td>
<td>3128.357</td>
<td>6.690</td>
<td>.004</td>
<td>.125</td>
</tr>
<tr>
<td>Error</td>
<td>36275.079</td>
<td>94 (HF77.574)</td>
<td>467.621</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS*Room</td>
<td>3240.370</td>
<td>4 (HF3.481)</td>
<td>930.903</td>
<td>6.442</td>
<td>.000</td>
<td>.121</td>
</tr>
<tr>
<td>Error</td>
<td>23642.519</td>
<td>188 (HF163.602)</td>
<td>144.513</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Room Condition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loudspeaker</td>
<td>2177.560</td>
<td>2 (HF1.917)</td>
<td>1136.126</td>
<td>11.531</td>
<td>.000</td>
<td>.197</td>
</tr>
<tr>
<td>Error</td>
<td>8875.995</td>
<td>94 (HF90.083)</td>
<td>98.532</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Room</td>
<td>88662.644</td>
<td>2 (HF1.748)</td>
<td>50726.928</td>
<td>133.135</td>
<td>.000</td>
<td>.739</td>
</tr>
<tr>
<td>Error</td>
<td>31300.245</td>
<td>94 (HF82.149)</td>
<td>381.020</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS*Room</td>
<td>1050.648</td>
<td>4 (HF4.000)</td>
<td>262.662</td>
<td>2.336</td>
<td>.057</td>
<td>.047</td>
</tr>
<tr>
<td>Error</td>
<td>21141.130</td>
<td>188 (HF163.602)</td>
<td>122.453</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3.16: Separate repeated measures of ANOVA analysis by condition. Data from experimental pages 7-9.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Type iii sum of squares</th>
<th>Df</th>
<th>Mean square</th>
<th>F</th>
<th>sig</th>
<th>Partial eta</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Loudspeaker</td>
<td>Condition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loudspeaker</td>
<td>21490.227</td>
<td>2 (HF 1.671)</td>
<td>12858.712</td>
<td>44.260</td>
<td>.000</td>
<td>.485</td>
</tr>
<tr>
<td>Error</td>
<td>22820.440</td>
<td>94 (HF 78.549)</td>
<td>290.524</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Room</td>
<td>4682.296</td>
<td>2 (HF 1.716)</td>
<td>2729.046</td>
<td>9.382</td>
<td>.000</td>
<td>.166</td>
</tr>
<tr>
<td>Error</td>
<td>23457.037</td>
<td>94 (HF 80.639)</td>
<td>290.889</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS*Room</td>
<td>2961.648</td>
<td>4 (HF 3.752)</td>
<td>789.381</td>
<td>4.728</td>
<td>.002</td>
<td>.091</td>
</tr>
<tr>
<td>Error</td>
<td>29441.685</td>
<td>188 (HF 176.337)</td>
<td>166.962</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Room</td>
<td>Condition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loudspeaker</td>
<td>1022.352</td>
<td>2 (HF 1.653)</td>
<td>618.362</td>
<td>5.478</td>
<td>.009</td>
<td>.104</td>
</tr>
<tr>
<td>Error</td>
<td>8772.093</td>
<td>94 (HF 77.706)</td>
<td>112.888</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Room</td>
<td>108232.171</td>
<td>2 (HF 1.699)</td>
<td>63719.875</td>
<td>138.328</td>
<td>.000</td>
<td>.746</td>
</tr>
<tr>
<td>Error</td>
<td>36774.273</td>
<td>94 (HF 79.832)</td>
<td>469.643</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS*Room</td>
<td>1849.718</td>
<td>4 (HF 3.601)</td>
<td>513.708</td>
<td>5.852</td>
<td>.000</td>
<td>.111</td>
</tr>
<tr>
<td>Error</td>
<td>14855.838</td>
<td>188 (HF 169.234)</td>
<td>87.783</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 3.17: Single repeated measures of ANOVA analysis, both conditions.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Type iii sum of squares</th>
<th>Df</th>
<th>Mean square</th>
<th>F</th>
<th>sig</th>
<th>Partial eta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loudspeaker</td>
<td>31413.197</td>
<td>2 (HF 2.000)</td>
<td>15706.598</td>
<td>84.677</td>
<td>.000</td>
<td>.372</td>
</tr>
<tr>
<td>Error</td>
<td>53049.470</td>
<td>286 (HF 296.000)</td>
<td>185.488</td>
<td>.641</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Room</td>
<td>211526.725</td>
<td>2 (HF 1.711)</td>
<td>123949.868</td>
<td>254.793</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>118716.609</td>
<td>286 (HF 244.628)</td>
<td>465.925</td>
<td>.002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>85.587</td>
<td>1 (HF 1.000)</td>
<td>85.587</td>
<td>.262</td>
<td>.609</td>
<td>.002</td>
</tr>
<tr>
<td>Error</td>
<td>46026.247</td>
<td>143 (HF 143.000)</td>
<td>320.072</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Room*LS</td>
<td>10583.106</td>
<td>4 (HF 3.523)</td>
<td>3004.271</td>
<td>19.109</td>
<td>.000</td>
<td>.118</td>
</tr>
<tr>
<td>Error</td>
<td>78950.894</td>
<td>572 (HF 503.714)</td>
<td>156.728</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS*Condition</td>
<td>19562.924</td>
<td>2 (HF 1.898)</td>
<td>10336.202</td>
<td>58.989</td>
<td>.000</td>
<td>.292</td>
</tr>
<tr>
<td>Error</td>
<td>47421.410</td>
<td>143 (HF 270.650)</td>
<td>322.24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Room* LS* Condition</td>
<td>97676.322</td>
<td>2 (HF 1.568)</td>
<td>62346.202</td>
<td>187.216</td>
<td>.000</td>
<td>.567</td>
</tr>
<tr>
<td>Error</td>
<td>74007.245</td>
<td>572 (HF 224.227)</td>
<td>332.732</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Room* Condition* LS* Room</td>
<td>72974.1</td>
<td>5 (HF 3.875)</td>
<td>188.296</td>
<td>1.728</td>
<td>.144</td>
<td>.012</td>
</tr>
<tr>
<td>Error</td>
<td>60372.926</td>
<td>572 (HF 554.195)</td>
<td>108.938</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The loudspeaker*condition interaction and the room*condition interaction are both significant (loudspeaker*condition: $F = 58.989$, $p < .001$, partial eta .292; room*condition: $F = 187.216$, $p < .001$, partial eta .567). The loudspeaker main effect for each condition is plotted in Figure 3.12 (A). The room main effect for each condition is plotted in Figure 3.12 (B). As was shown in Figure 3.12 the interaction is in the correct direction to show compensation for both factors.

**Hypothesis 1d**

Tables 3.14-3.16 show significant loudspeaker*room interactions in both conditions. These may be evidence against compensation or show that compensation has not occurred at all levels of the indirect factor. Loudspeaker*room effects are expected to reduce alongside main effects where compensation has occurred. Because of the experimental set-up it was not possible to test for a reduction in room*loudspeaker interactions between conditions. However, it is possible to examine the size of the interactions with large loudspeaker*room interactions indicating incomplete compensation.

In the data in Figure 3.8 it is apparent that LS1 is particularly disliked in R2, and in Figure 3.10 it is apparent that R2 is particularly liked when listening to LS3. It may be that these loudspeaker and room combinations are particularly unattractive/attractive because of the way the loudspeaker and room acoustics combine, in which case they show a continuing effect of the indirect factor. However, the explanation for these interactions may not be so straightforward. For example it is possible (but by no means clear) that the interaction seen in Figure 3.8 may represent an increase in spread of ratings for R2, in which case it is possible that this increase in spread is caused by there being less effect of the indirect factor (room) at this level of the indirect factor (R2) allowing loudspeaker differences to be heard more clearly, rather than the room having an effect to change the sound of LS1. Again a fully intermixed condition could be used to test this in future. It is clear from the partial eta measures and Figures 3.8 and 3.10 that loudspeaker*room interactions, although statistically significant, are small. The absence of large room*loudspeaker interactions is confirmation that the indirect factor in each condition is not having a large effect on ratings.
Hypothesis 1e

A gradual reduction in the effect of the indirect factor across each block may be expected if compensation is the result of longer continuous listening, and compensation occurs over the course of seconds and minutes. This sort of compensation suggests a fatiguing process such as that seen in loudness adaptation (Hood 1950) whereby the listener becomes less sensitive to the loudness of a pure tone over the periods of seconds and minutes. A similar fatiguing process may occur for timbre, where the listener becomes less sensitive to the timbre of a stimulus that is heard continuously over seconds to minutes. The effect of the indirect factor is expected to be smaller with ratings that occur later in the experimental block if compensation due to the length of continuous listening to the indirect factor has occurred. The listeners make their first ratings on page 1, having listened to the indirect factor continuously for approximately 25 s and they make their last ratings for the indirect factor on page 9, having listened to it continuously for approximately 4 minutes (see Table 3.11 for average times spent listening within each block). The analysis in Tables 3.15-3.16 was compared for evidence of an indirect factor*experimental page interaction. The difference between rooms when compared indirectly is actually larger with increased experimental page. It is possible that this result shows that listeners became more practised at hearing the difference between rooms as the block continued, instead of becoming less sensitive. The difference between loudspeakers, when compared indirectly, is smaller with continuous listening but not significantly so. There is no consistent reduction in loudspeaker*room interactions in the expected direction with later experiment pages, as would be expected if the indirect factor becomes smaller with continuous listening. These results show that variation in the indirect factor does not decrease with continuous listening between approximately 25 s-4 minutes. However, these results do not show that longer continuous listening does not cause compensation. It may be that most of the compensation for the indirect factor occurs before the time that listeners make their first ratings (before approximately 25 s). A loudness adaptation-style fatiguing process can occur within 25 s (Egan 1955, Hood 1950). Compensation occurring under 25 s could not be measured in this test because no measurement was taken earlier than this. It is also possible that continuous listening beyond 4 minutes may cause further compensation but a reduction with continuous
listening over approximately 4 minutes was not measured here. These results may be considered a preliminary investigation of the effect of longer listening on compensation as it was not a main aim of this experiment to measure this effect and the effect could only be measured under limited circumstances. If future studies are run which specifically aim to investigate the effect of longer listening on compensation, it should be possible to measure compensation under 25 seconds, where compensation might be expected to occur most readily and the time course of listening should be more tightly controlled.

3.11 Experiment 2: Summary and discussion

Experiment 1 revealed the need for further confirmation of compensation for loudspeakers and rooms in real-world listening. The current experiment aimed to replicate Olive et al.’s (1995) study using materials and conditions that are more favourable to compensation than those used in Experiment 1. It was necessary to ensure that perceived differences between stimuli were large enough when compared under optimal conditions (short listening time courses) so that any reduction in difference that may occur due to compensation could be measured. This was done by using a speech programme item which was thought to be more revealing of the loudspeaker and room differences and likely to reduce the possibility of floor effects seen in Experiment 1. The loudspeakers and rooms were not changed to more extreme examples to ensure they remained representative of real-world channels but compensation was encouraged by increasing the length of the time gaps between stimulus presentations from 2 to 10 minutes for the indirectly compared stimuli, and by increasing the length of continuous listening to the indirectly compared stimuli by a factor of 9 compared to Experiment 1. If compensation is dependent on the longer listening time courses (longer time gaps and/or longer continuous listening) then this should increase the compensation effect.

The experiment showed statistically significant compensation for the loudspeaker and room factor at the p<.05 level. Alongside the work of Olive et al. and Experiment 1, compensation for these channels is now confirmed. Further, the use
of new material in this study has extended the validity of Olive et al.'s results: compensation is shown to occur with speech (this study) as well as non-speech (Olive et al.'s study and Experiment 1) and with new transmission channels. The evidence of real-world compensation is now similar for speech and music (at least 1 study confirming compensation with each) and all 3 studies together provide good evidence that compensation occurs with this paradigm. However, it still may be revealed that the compensation seen here is different with different programme items. It appears that the changes between this experiment and Experiment 1 may have brought about larger compensation. In particular the speech programme item may have both increased perceived difference between loudspeakers and rooms when compared directly (reducing floor effects) and it may have increased the compensation effect due compensation occurring more readily when listening to speech. It might be useful to conduct an experiment that investigates both speech and non-speech stimuli in a single test in future work, so that differences in the size of compensation between programme items can be compared. However, it is not necessary to conduct this in order to answer the research questions in this thesis. Additionally, if time course is important to listening then increasing this compared to Experiment 1 may have brought about the larger compensation effect. Increasing the number of repeated ratings increases statistical power and makes significant compensation more likely to be observed due to increased accuracy in measurement but increasing the repeated ratings also caused longer continuous listening to the channel and this may have brought about larger compensation. Time gaps were also increased in length.

In this experiment significant compensation was observed for the loudspeaker factor and main effects were significant with both direct and indirect comparisons showing that floor effects did not prevent significant compensation being measured this time. However, it should be noted that only some of the channels tested showed a movement indicative of compensation (LS3). A floor effect was in fact experienced for LS2 and LS1, which were not different even with direct comparisons. If future tests are run using this paradigm, then ideally these should not just require a significant main effect, but a significant difference between all the channels being compared (i.e. between LS1 and LS2, LS1 and LS3, and LS2 and LS3), so that a movement indicative of compensation can be measured for all channels in the test, strengthening the validity of results. The 3
3.11 EXPERIMENT 2: SUMMARY AND DISCUSSION

Rooms tested were all different with direct comparisons and compensation was observed for each of these, contributing to a larger and more valid compensation effect being observed for the room factor.

The current experiment confirms that compensation decreases perceptions of channel differences. Descriptive statistics show that the magnitude of compensation found in this experiment was larger than in Experiment 1 for both loudspeakers and rooms. Experiment 2 used a more valid method of quantifying the magnitude of compensation compared to Olive et al.’s (1995) experiment and Experiment 1: compensation was described using the difference in the range of loudspeaker/room ratings with direct versus indirect comparisons using rating scale values. There was movement of 10 points between direct and indirect conditions for the loudspeaker factor, which is a ‘small to ‘moderate’ reduction in channel effect according to rating scale values. There was movement of 29 points between direct and indirect conditions for the room factor which was a ‘large’ reduction of effect according to rating scale values. Examining changes using the preference rating scale provides a perceptual measure of effect size with good face validity and this measure can be easily compared across studies using this rating scale. It is not clear why compensation for the loudspeaker factor was smaller than for room factor in this study but this may be to do with the underlying differences between rooms being larger than was the case for loudspeakers. It cannot be said that that floor effects prevented compensation for either factor in this study but it may be that where variation is larger to begin with more compensation might occur. Alternatively, the mechanisms of compensation may be different for these factors.

The fact that effects of rooms and loudspeakers remained significant with indirect comparisons and the presence of room*loudspeaker interactions indicates that compensation was not complete: the indirect factor still had an effect on ratings. Therefore, in all cases the channel is perceived. It is not clear whether the size of compensation seen is sufficient to ensure timbral constancy and accurate identification of sounds. It is possible that only a small amount of compensation is necessary for constancy and also that other (non-compensation) mechanisms may contribute to auditory object recognition in the face of spectral distortion in order to achieve constancy. It would be interesting to further examine how the magnitude of compensation relates to timbral constancy but this is not necessary to answer the
research questions set out in this thesis.

This experiment did not specifically examine the time course of the compensation, however, compensation was increased compared to Experiment 1 as expected. This increased compensation suggests that the longer time course of listening might cause compensation: the length of listening or time gaps may be important to causing compensation. The fact that compensation was reduced in Experiment 1 suggests that it may be that more than 2 minute gaps are needed for compensation. It may also be that the increased continuous listening in Experiment 2 (and in Olive’s test) brought about more compensation. However, the length of listening hypothesis is put into doubt, as no further decrease in the channel was observed when listening to the indirect factor continued across the experimental block. This appears to rule out a hypothesis that fatigue to timbre that occurs over seconds and minutes of listening causes the compensation seen here. However, it is noted that it was not possible to measure any compensation that occurred before 25 s (during which time the indirect factor had been heard continuously, but ratings not yet made). It is possible, and perhaps likely, that compensation is a reasonably quick acting process and would occur before 25 s. Contrary to expectation, there was less compensation with continuous listening across the experimental block for the room factor. This may suggest that the listener is learning the channel and becoming more sensitive, rather than less sensitive, over time. A learning of the channel does not necessarily imply that compensation does not also occur because it is possible that an automatic channel compensation occurs with some listening to the channel (this may help a listener in a new environment) but that with further listening this compensation effect is reduced somewhat as the listener learns the channel. This may explain why experienced listeners can hear channel differences. It can be concluded that longer listening between approximately 25 s and 4 minutes is not likely to be relevant to the compensation seen in Olive et al.’s (1995) study and Experiments 1 and 2, but listening outside of this time period may be. Further research would be needed to examine the effect of longer listening on compensation. Time gaps are also thought to be relevant and these were Olive et al.’s (1995) primary explanation for the compensation seen in their studies. It might be that reduced compensation occurred in Experiment 1 because of the 2 minute rather than 10 minute gaps used in that study. The current study therefore indicates that time gaps over 2 minutes may
be necessary for significant compensation via a time-gap sensitive mechanism to occur. As time gaps were offered by Olive et al. (1995) as the primary explanation for their results and as the current study casts doubt on the role of longer continuous listening real-world compensation compensation, the next experiment in this this thesis will proceed to investigate the effect of time gap on compensation further. This experiment will also confirm that the results in Olive et al.’s (1995) experiment and Experiments 1 and 2 are not an artefact of the experimental process and caused by the specific instructions given to listeners.

3.12 Experiment 2: Conclusion

Experiment 1 failed to show significant compensation. Experiment 2 aimed to replicate the experiment by Olive et al. (1995) and show significant compensation for loudspeakers and rooms when conditions of the experiment were more favourable to observing compensation. Significant compensation for loudspeakers and rooms was seen in Experiment 2. The magnitude of compensation was quantified in a meaningful way using the timbral rating scale. Compensation for the room factor was large (29 preference points). Compensation for loudspeaker factor was smaller but still ‘moderate’ (10 preference points). A speech programme item was used in this test. There is now evidence that compensation occurs with both speech and musical programme items. Now that compensation has been confirmed to occur with this paradigm it can be concluded that paradigm is useful for examining compensation further. Future tests of compensation mechanisms can be run with this paradigm.

The significant compensation in the current study but not in Experiment 1 appears to be due to the conditions being more favourable to observing compensation. The change in conditions hints at the mechanisms that may be behind compensation. The length of listening with direct comparisons was increased compared to Experiment 1 up to approximately 4 minutes (which is is likely to be more similar to that used in Olive et al.’s original study). Larger compensation in this test, compared to Experiment 1, may be because of this longer continuous listening to the channel. However, Experiment 2 shows that compensation is not likely to occur because of a fatiguing process acting
between approximately 25 s and 4 minutes of listening. In fact it was shown that the listener can become more sensitive to channel effects over this time. This does not necessarily put compensation into question as compensation was still apparent and this is more likely to occur over a shorter time frame e.g. under 25 s (as is seen with loudness adaptation). It is therefore acknowledged that while listening between 25 s and 4 minutes may not be central to the compensation seen, length of listening under 25 s may be a factor. This test also increased the time gap between hearing different channels back to that used in Olive et al.’s original study. Larger compensation in this test compared to Experiment 1 may be due to the longer time between hearing different channels. Further research is now necessary to establish the mechanisms of real-world compensation. Olive et al. (1995) appear to prefer the time gap explanation for compensation. Therefore, this explanation will be investigated in Experiment 3 in the next section (Section 3.13). Another hypothesis, that compensation seen in real-world tests is caused by the task instructions, is also tested.

3.13 Experiment 3: Research aims and rationale

Experiment 3 tests the mechanisms of compensation for loudspeakers and rooms seen in Experiments 1 and 2 and in Olive et al.’s (1995) study. The current experiment involves running an experiment under the same conditions as Experiment 2 but with experimental manipulations necessary to establish the mechanisms behind compensation. The primary explanation for real-world compensation given by Olive et al. was the time between comparisons when stimuli are compared indirectly. The current study will test whether time between comparisons (hereafter time gaps) is a factor in compensation. The role of time-gap sensitive mechanisms will be tested by removing the time gap between indirectly compared stimuli and examining the extent to which compensation reduces. If compensation is eliminated then a time-gap sensitive mechanism is the sole cause of the compensation seen in Olive et al. (1995) and Experiments 1 and 2. Mechanisms that are dependent on the time between listening to different stimuli may explain compensation. If any compensation remains this must be due to non-time-gap sensitive mechanisms— possibly including a mechanism that requires longer continuous listening (outside 25 s-4 minutes).
It must also be established that the compensation seen in Olive et al.’s study and Experiments 1 and 2 is not caused by task-specific effects which would not occur in a normal real-world listening scenario. A cause of the reduction in stimuli variation with indirect comparisons may be bias caused by the task set-up or task instructions. In the previous experiments listeners were explicitly instructed to compare differences between stimuli on each experimental page (the direct factor), but were not explicitly told that global changes across blocks would occur and that these should also be reflected in ratings. If the listeners thought that the purpose of the task was to measure differences occurring on each page and that any changes across pages were irrelevant, they may have intentionally recalibrated their rating scales with each new page to use 50 points as a midpoint for preference. This bias towards rating with regard to the local context would result in differences across blocks (differences due to the indirect factor) being removed from ratings. Such a bias may be less likely to occur in real-world listening, where listeners are likely to listen more globally. Instructions to listen globally for changes across the test and report these in ratings would be expected to reduce compensation if this sort of intentional response bias was present. If compensation remains after this bias is removed then it can be concluded that compensation seen is genuine rather than a result of an experimental design artefact that resulted in listeners ignoring changes across blocks when making ratings.

The research in Experiment 3 therefore aims to answer Research Question 1b. In order to answer this question, this chapter will also aim to answer the following chapter questions:

Question 3.2: Is this compensation an artefact of the experimental process or is it likely to be a genuine listening phenomenon?

Question 3.3: What type of perceptual mechanisms might be behind this compensation?

3.14 Experiment 3: Hypotheses

It was hypothesised that the elimination of the time gap would lead to reduced compensation if a time-gap sensitive mechanism causes compensation. Any compensation
remaining after this manipulation suggests a non-time-gap sensitive cause. It was also hypothesised that compensation would be reduced or eliminated in a condition where instructions to rate globally are given (the instructions condition) compared to a condition where no specific instructions are given (no-instructions condition). This manipulation is expected to remove any possibility that the task set-up or instructions, as administered in previous tests, bias listeners to make local rather than global ratings. If compensation remains as large in the instructions as in the no-instructions, then it can be concluded that compensation is a genuine listening phenomenon and that the compensation seen in Olive et al’s work and Experiments 1 and 2 was not due to the task setup/instructions.

### 3.15 Experiment 3: Participants and materials

Sixteen listeners none of whom had participated in Experiment 1 and 2 were selected using the same criteria as in Experiments 1 and 2 (see Section 3.2.1). All materials and the procedure were the same as Experiment 2 except for the removal of break phases between blocks and additional instruction to make global ratings for the instruction group (n=8). All stimuli were the same as those used in Experiment 2 (see Section 3.8).

### 3.16 Experiment 3: Procedure

The procedure is the same as in Experiment 2 (see Section 3.9). However, in this experiment the 10 minute break periods within the indirect comparison condition were removed so that listeners heard the changes between stimuli side-by-side with no time between them like when listening to the directly compared stimuli. The results of this condition are then compared to the the 10 minute gap indirect condition (in Experiment 2). Further, half of the listeners were given specific instructions to make ‘global ratings reflecting differences in stimuli across the test as well as on a single page’ (the instruction condition). Further explanation of this request was given if needed. The remaining listeners were given the same instructions as in Experiments 1 and 2.
(no instruction condition), which was simply to ‘rate timbral differences between the stimuli’ on each experimental page. These non-specific instructions appear to be the same as those used by Olive et al. (1995). The amount of compensation was compared between participants who had received the specific global instructions and those who had not.

3.17 Experiment 3: Results

This section describes the descriptive measures (Section 3.17.1) and statistical analysis to examine compensation (Section 3.17.2). Data was checked for extreme departures from normality using Q-Q plots. No data was removed because of unreliability. Where Mauchly’s test of sphericity indicated that the data was non-spherical Huynh-Feldt adjusted degrees of freedom were used to produce variance measures.

3.17.1 Descriptive measures

This section examines graphs and descriptive statistics with the aim of assessing whether the variation in loudspeakers and rooms is larger when direct comparisons are made compared to indirect comparisons (i.e. whether compensation occurs), whether compensation decreases when 0 minute time gaps are used compared to 10 minute time gaps, and whether compensation reduces with instructions to make global ratings compared to when listeners are not given specific instructions to rate globally.

The effect of time gap

As in Figures 3.8 and 3.10, Figure 3.13 (Top) shows the loudspeaker ratings when comparing loudspeakers directly (A) and indirectly (B) with 10 minute time gaps between indirect comparisons (the 10 minute condition). Different loudspeakers are displayed as different coloured lines. Figure 3.13 (Top) also shows room ratings when compared indirectly (C) and directly (D) in the 10 minute condition. Figure 3.13 (Bottom) shows the same results when there was no time gap between indirectly compared stimuli (0 minute condition). The values used to produce this graph can
be seen in Tables 3.18 and 3.19 and can be compared with those displayed in Tables 3.12 and 3.13.

Figure 3.13: Top: 10 minute condition: the effect of loudspeaker in the Loudspeaker condition (A) and Room condition (B) and the effect of room in the Loudspeaker condition (C) and the Room condition (D). Differences in loudspeakers are shown as differences between lines in A and B. Differences in rooms are show as differences between lines in C and D. Bottom: the equivalent data for the 0 minute condition.

As in previous experiments compensation is denoted by a reduction in variation between ratings for a factor when comparisons are made indirectly compared to directly. As noted previously this reduction is shown for loudspeakers when there are 10 minute...
3.17 EXPERIMENT 3: RESULTS

Table 3.18: Loudspeaker means within each room and loudspeaker mean variance within each room, in the 0 minute time gap condition.

<table>
<thead>
<tr>
<th>Loudspeaker</th>
<th>Condition</th>
<th>Room</th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
</tr>
</thead>
<tbody>
<tr>
<td>LS1 Mean</td>
<td>Direct</td>
<td>LS</td>
<td>42.49</td>
<td>54.89</td>
<td>58.86</td>
</tr>
<tr>
<td>LS2 Mean</td>
<td>Direct</td>
<td></td>
<td>40.31</td>
<td>58.51</td>
<td>56.50</td>
</tr>
<tr>
<td>LS3 Mean</td>
<td>Direct</td>
<td></td>
<td>45.78</td>
<td>68.53</td>
<td>68.74</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>Direct</td>
<td>LS</td>
<td>42.86</td>
<td>60.64</td>
<td>61.37</td>
</tr>
<tr>
<td><strong>Variance</strong></td>
<td>Direct</td>
<td></td>
<td>7.59</td>
<td>49.91</td>
<td>42.14</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>Direct</td>
<td></td>
<td>2.75</td>
<td>7.06</td>
<td>6.49</td>
</tr>
<tr>
<td>LS1 Mean</td>
<td>Indirect</td>
<td>Room</td>
<td>31.32</td>
<td>65.22</td>
<td>70.39</td>
</tr>
<tr>
<td>LS2 Mean</td>
<td>Indirect</td>
<td></td>
<td>31.21</td>
<td>70.11</td>
<td>69.19</td>
</tr>
<tr>
<td>LS3 Mean</td>
<td>Indirect</td>
<td></td>
<td>30.57</td>
<td>69.90</td>
<td>69.54</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>Indirect</td>
<td>Room</td>
<td>31.03</td>
<td>68.41</td>
<td>69.71</td>
</tr>
<tr>
<td><strong>Variance</strong></td>
<td>Indirect</td>
<td></td>
<td>0.16</td>
<td>7.64</td>
<td>0.38</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>Indirect</td>
<td></td>
<td>0.40</td>
<td>2.76</td>
<td>0.61</td>
</tr>
<tr>
<td><strong>SD Ratio</strong></td>
<td></td>
<td></td>
<td>6.81</td>
<td>2.56</td>
<td>10.56</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Room</th>
<th>Condition</th>
<th>Loudspeaker</th>
<th>LS1</th>
<th>LS2</th>
<th>LS3</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1 mean</td>
<td>Room</td>
<td>Direct</td>
<td>31.32</td>
<td>31.21</td>
<td>30.57</td>
</tr>
<tr>
<td>R2 mean</td>
<td>Room</td>
<td>Direct</td>
<td>65.22</td>
<td>70.11</td>
<td>69.90</td>
</tr>
<tr>
<td>R3 mean</td>
<td>Room</td>
<td>Direct</td>
<td>70.39</td>
<td>69.19</td>
<td>69.54</td>
</tr>
<tr>
<td><strong>mean</strong></td>
<td>Room</td>
<td>Direct</td>
<td>55.64</td>
<td>56.84</td>
<td>56.67</td>
</tr>
<tr>
<td><strong>Variance</strong></td>
<td>Room</td>
<td>Direct</td>
<td>450.42</td>
<td>492.87</td>
<td>511.01</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>Room</td>
<td>Direct</td>
<td>21.22</td>
<td>22.20</td>
<td>22.61</td>
</tr>
<tr>
<td>R1 mean</td>
<td>Indirect</td>
<td>LS</td>
<td>42.49</td>
<td>40.31</td>
<td>45.78</td>
</tr>
<tr>
<td>R2 mean</td>
<td>Indirect</td>
<td></td>
<td>54.89</td>
<td>58.51</td>
<td>68.53</td>
</tr>
<tr>
<td>R3 mean</td>
<td>Indirect</td>
<td></td>
<td>58.86</td>
<td>56.50</td>
<td>68.74</td>
</tr>
<tr>
<td><strong>mean</strong></td>
<td>Indirect</td>
<td>LS</td>
<td>52.08</td>
<td>51.77</td>
<td>61.01</td>
</tr>
<tr>
<td><strong>Variance</strong></td>
<td>Indirect</td>
<td>LS</td>
<td>72.96</td>
<td>99.64</td>
<td>174.12</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>Indirect</td>
<td>LS</td>
<td>8.54</td>
<td>9.98</td>
<td>13.20</td>
</tr>
<tr>
<td><strong>SD Ratio</strong></td>
<td></td>
<td></td>
<td>2.48</td>
<td>2.22</td>
<td>1.71</td>
</tr>
</tbody>
</table>

Table 3.19: Room means for each loudspeaker and room mean variance for each loudspeaker. In the 0 minute time gap condition.
gaps between indirectly compared stimuli (i.e. the loudspeaker ratings in figure 3.13 (Top) are reduced in variation in panel B compared to A). This reduction is also shown for rooms (ratings in Figure 3.13 (Top) are reduced in panel C compared to panel D). From looking at Figure 3.13 (Bottom) it is apparent that the same compensation occurs when there are no time gaps between indirect comparisons (the 0 minute condition). The results presented here therefore appear to provide another demonstration of compensation for loudspeakers and rooms. It should also be noted that order of preference is similar across 0 and 10 minute conditions with LS1 and R1 consistently the worst performers and there is generally a lack of significant disordinal interactions, which shows stability in the order of preference between direct and indirect comparisons and between 0 minute and 10 minute conditions, as is expected by the compensation hypothesis.

Because compensation is seen with 0 minute time gaps, the removal of time gaps cannot be said to eliminate compensation. Time gap is therefore not the sole cause of compensation. However, a reduction in compensation shows that time gap is partly a cause. For the room factor there is an increased spread in ratings towards that observed with direct comparisons. Figure 3.13 (Bottom, Panel C), shows that there is less reduction in room variance with indirect comparisons in the 0 minute condition compared to the 10 minute condition (Figure 3.13 (Top, Panel C)). If significant, this trend is suggestive of the time gap between indirect comparisons causing compensation. However, the loudspeaker factor does not appear to show increased variation with indirect comparisons in the 0 minute time gap condition compared to the 10 minute condition (i.e. there does not appear to be a significant expansion of ratings in Figure 3.13 (Bottom, panel B) compared to Figure 3.13 (Top, Panel B)). Therefore, compensation has not decreased with the removal of time gaps for the loudspeaker factor suggesting that there are other causes of this compensation.

The pattern of increased variation with indirect comparisons in the 0 minute time gap condition is suggestive of a role of the time gap in compensation (for rooms at least). However, the removal of time gap also appears to affect direct comparisons and this means that an analysis that only involves the examination of variation in the indirect comparison condition is incomplete and possibly misleading. To get a better idea of changes in compensation with time gap, and of the magnitude of these changes,
changes in the range of preference ratings within each condition must be considered. Main effects are analysed to examine effects more closely.

**Figure 3.14:** Top) Main effect of loudspeaker in the Loudspeaker condition (A) and in the room condition (B) in the 10 minute (Red) and 0 minute conditions (Blue). Bottom) Main effect of room in the Loudspeaker condition (C) and Room condition (D) for the 10 minute (Red) and 0 minute conditions (Blue).

**Loudspeaker factor**

For the loudspeaker factor in the 10 minute condition the difference between the most and least preferred loudspeaker when directly compared (Figure 3.14 (Top, panel A)) was 13 points (averaged across rooms) compared to 3 points when indirectly compared in Figure 3.14 (Top, panel B)—showing 10 points compensation, which is moderate according to rating scale values. In the 0 minute gap condition the difference between the most preferred and the least preferred loudspeaker was 9 points when directly compared (Figure 3.14 Top, panel A), compared to 1 point when indirectly compared (Figure 3.14 (Top, panel B)), showing 8 points compensation. The reduction in compensation of 2 points suggests that there is less compensation in the 0 minute condition. However, it can be seen that the effect is largely not in the correct direction to show that time gap explains compensation. The removal of time gaps decreases range in the direct comparison condition more than it increases the range in the indirect condition. Range reduced by 4 points when loudspeakers were compared directly in the 0 minute condition, compared to the 10 minute condition (Figure 3.14 (Top, Panel A)) and range is decreased by 2 points with indirect comparisons, with
0 minute condition compared to the 10 minute condition (Figure 3.14 (Top, Panel B)). Therefore, the reduced compensation seen in the 0 minute condition is more because of listeners’ tendency to hear loudspeakers as more similar when rating directly rather than a tendency to perceive loudspeaker differences as larger when rating indirectly (i.e. reduced compensation). Therefore, it appears that time gaps do not explain compensation for the loudspeaker factor. It is not clear why removing the time gap causes slightly decreased variation in ratings when rating directly. If significant, this decrease needs to be explained. It is possible that because there is now a larger difference perceived between the indirectly compared stimuli (rooms) this requires more space on the rating scale so the listener makes more room for this by reducing the space taken up but the directly compared factor. Further, it is possible that because in this test the room differences are large when compared directly there is no further space on the rating scale for loudspeaker differences to increase when compared indirectly. These issues may explain the lack of effect of time gap for the loudspeaker factor. It is also apparent that there is notable compensation remaining after the time gaps are removed (8 points), which needs to be explained. Non-time-gap sensitive factors must be behind this remaining compensation.

**Room factor**

For the room factor the time gap does appear to be a cause of the compensation. In the 10 minute time gap condition, the difference between the most preferred room and the least preferred room was 36 points when directly compared (Figure 3.14 (Bottom, Panel D)), compared with 7 points when indirectly compared (Figure 3.14 (Bottom, panel C)), showing a ‘large’ compensation of 29 points. In the 0 minute time gap condition this difference was 39 points when directly compared (Figure 3.14 (Bottom, panel D)) and 18 points when indirectly compared (Figure 3.14 (Bottom, panel C)), showing a ‘moderate-large’ compensation of 21 points. The reduction in compensation between the 0 and 10 minute conditions is approaching moderate (9 points). Unlike for the loudspeaker factor, this change is in the direction that would be expected if time gaps cause compensation. The majority of the difference is due to the increase in range from 7 points (10 minute condition) to 18 points (0 minute condition) when making indirect comparisons (a 11 point increase in range). However, the room range between direct comparison conditions across time gap conditions has also increased slightly (a
3.17 EXPERIMENT 3: RESULTS

3 point increase in range with 0 minute time gap). It appears that the removal of time
gaps not only aided listeners in hearing differences between rooms when compared
indirectly, which is expected by the compensation hypothesis, but it also appears to
have helped listeners make comparisons directly to some extent. It is not surprising
that a reduction in the time gap may aid direct comparisons as well as indirect as being
able to hear the whole range of stimuli closer in time may increase sensitivity to all
differences in the test. Compensation for the room factor may be said to be explained,
in part, by the time gaps between stimuli.

Effect of instructions

Data obtained when listeners were given instructions to make global ratings was
compared with data obtained where listeners were under no specific instructions (as in
the experiments by Olive et al. (1995) and Experiments 1 and 2).

As in Figure 3.13, Figure 3.15 (Top) shows the loudspeaker ratings when comparing
loudspeakers directly (A) and indirectly (B) without instructions (the no instruction
condition). Different loudspeakers are displayed as different coloured lines. Figure 3.15
(Top) also shows room ratings when compared indirectly (C) and directly (D) in the
no instructions condition. Different rooms are displayed as different coloured lines.
Figure 3.15 (Bottom) shows the same results when instructions to rate globally were
given (the instructions condition). A figure with the same data but only displaying the
main effects is provided (see Figure 3.16). For accuracy and simplicity only this main
effects figure is referred to in the description of the results in the text below.

If instructions are the sole cause of the compensation seen previously it is expected
that variation with indirect comparisons will increase to the level seen with direct
comparisons. Figure 3.16 (Top) shows the main effect for the loudspeaker factor when
compared directly (panel A) and indirectly (panel B) with and without instructions.
A compensation effect can be seen whereby there is a decrease in variation with
indirect comparisons both with and without instructions. Without instructions the
range of loudspeaker ratings decreases by 8 points in panel B compared to A (‘small’
compensation). With instructions, the range of loudspeaker ratings decreases by 4
points (negligible compensation). The fact that compensation for the loudspeaker
3.17 EXPERIMENT 3: RESULTS

Figure 3.15: Top) No Instructions condition: the effect of loudspeaker in the Loudspeaker condition (A) and the Room condition (B) and the effect of room in the Loudspeaker condition (C) and Room condition (D). Bottom) the same data for the Instructions condition. Differences in loudspeakers are shown as differences between lines in A and B. Differences in rooms are shown as differences between lines in C and D. Bottom: the same data for the Instructions condition.

A factor is reduced by 4 points with instructions to make global ratings is suggestive of instructions causing compensation. However, closer examination reveals that this reduction in compensation occurs largely because variation in loudspeakers has reduced with direct comparisons rather than any notable increase in variation
3.17 EXPERIMENT 3: RESULTS

Figure 3.16: Loudspeaker main effects when compared with global instructions and no specific instructions, in the Loudspeaker (Panel A) and Room (panel B) conditions. Bottom): Room main effects in the Room (Panel D) and Loudspeaker (Panel C) conditions

with indirect comparisons. The range with direct comparisons reduced from 9 to 6 points with instructions, the range with indirect comparisons increased from 1 to 2 points. Therefore, instructions appear to mainly cause a tendency for listeners to be less sensitive when rating loudspeakers directly, which is not expected under the compensation hypothesis. There is also a small tendency toward increased sensitivity when rating indirectly, which is indicative of instructions causing compensation, but this increase of 1 point is negligible according to rating scale values. As was explained, with the removal of time gaps, it is possible that where the indirect factor has become more prominent a decrease in ratings for the direct factor would be expected. However, it can be seen in Figure 3.16 (Bottom, Panel C) that indirect the factor (the room) is not more prominent with instructions compared to without, so this reduction remains unexplained. It can be concluded that instructions do not cause compensation for the loudspeaker factor: even when listeners are instructed to make global ratings they remain insensitive to differences between stimuli when compared indirectly. In fact instructions result in a tendency to be less sensitive when making direct comparisons. Factors other than instructions appear to explain compensation for the loudspeaker factor.
3.17 EXPERIMENT 3: RESULTS

The room factor also shows no effect of instructions. Importantly, there is no increased variation with indirect comparisons where instructions are given to make global ratings compared to when no instructions are given (Figure 3.16 (Bottom, Panel C)). Without instructions there is a 39 point difference between the least and most preferred room when rated directly, and a 18 point difference when rated indirectly (a 21 point ‘large’ compensation). When instructions are used, there is a 32 point difference between the least and most preferred room when rated directly and there is a difference of 19 points, when rated indirectly (13 point ‘moderate’ compensation). There is an 8 point reduction in room range with instructions indicating an 8 point reduction in compensation, but this is driven by the fact that variation with direct comparisons is reduced with instructions (Figure 3.16 (Bottom, Panel D)): 32 points with instructions, compared to 39 points without instructions (7 points) rather than an increase when rated indirectly of 19 points versus 18 points (1 point). This decrease in variation in the instruction conditions when rooms are compared directly suggests that the listener appears to find it more difficult to observe differences when instructed to do so, when comparing directly. It is not clear why a decrease in direct ratings would occur. It may be that the listeners are making room on their rating scales for differences across the test that never materialise. Even though they are now attending to differences between indirectly compared stimuli they do not hear large differences.

Overall the results appear to show that encouraging listeners to listen for changes across the whole experiment (i.e. changes due to the indirect factor) by giving instructions to make global ratings was not effective in increasing listeners’ tendency to report differences between indirectly compared stimuli. Instead these instructions appeared to result in listeners having a slight tendency to be less sensitive to differences between directly compared stimuli, which is not expected if instructions explain compensation. Increased conservatism in reporting differences for the indirect factor, due to making space on the rating scale for the other factor, does not appear to explain the results in this case. However, there may be increased conservatism in rating the direct factor due to expected large differences across the test which do not materialise.
3.17 EXPERIMENT 3: RESULTS

3.17.2 Statistical analysis

The pattern of results described in the previous section was analysed to determine whether the trends described were statistically significant.

Compensation

The results from the 0 minute time gap condition were analysed with separate Repeated-Measures Analysis of Variance (RMANOVA) analyses for the loudspeaker and the room condition. In the loudspeaker condition a significant effect of loudspeaker occurred \( F = 39.303, p < .001 \), but not in the room condition \( F = 1.391, p = .252 \). The loudspeaker-by-condition (direct versus indirect) interaction was significant \( F = 25.837, p < .001 \). The effect of room was significant in the room condition \( F = 330.830, p < .001 \) and in the loudspeaker condition \( F = 62.258, p < .001 \). The room-by-condition (direct versus indirect) interaction was significant \( F = 46.427, p < .001 \). As discussed in the descriptive statistics section (3.17.1) the direction of these interactions is as expected so Experiment 3 has provided further evidence of significant compensation for loudspeakers and rooms additional that seen in Olive et al.’s (1995) experiment and Experiments 1 and 2.

The effect of time gap on compensation

To test whether compensation was significantly reduced with 0 versus 10 minute time gaps for both factors a single mixed ANOVA was conducted. The loudspeaker-by-condition (indirect versus direct)-by-time gap (0 versus 10 minutes) interaction was significant but small \( F = 6.430, p = .002 \) partial eta squared = .021. As was described in the previous section the effect is largely in the opposite direction to that expected if time gap causes compensation. Therefore, this significant interaction may not show that time gaps caused the compensation for the loudspeaker factor. However, the results do appear to show that ratings for loudspeakers in Experiment 3 are significantly smaller when compared directly, and as was discussed in the previous section, it is possible that the now larger room ratings have both prevented a reduction in compensation with time gaps from being revealed and caused a reduction in loudspeaker variation when rated directly, due to rating scale limitations. Further tests could be run to confirm that, as hypothesised here, time gap does effect loudspeaker ratings.
where the rating scale is expanded and ceiling effects are prevented. The 8 points worth of compensation remaining after time gap explanation is eliminated also needs to be explained. This must be caused by a non-time-gap sensitive mechanisms.

To test whether compensation was significantly reduced with 0-minute compared to 10-minute time gaps for the room factor, a single mixed ANOVA was conducted. The room-by-condition (direct versus indirect)-by-time gap (0 versus 10 minutes) interaction was highly significant ($F = 16.497, p = .000$ partial eta squared $= .055$). This interaction was in the direction expected to show a contribution of time gap to compensation. Both the partial eta and the extent of movement of 9 points shows an almost ‘moderate’ reduction in compensation caused by removing the time gap. Therefore, time gaps can be concluded to play a moderate role in compensation for rooms. There is 20 points worth of compensation that remains to be explained. Other, non-time-gap sensitive, causes must also contribute to compensation.

In light of these observations it can be concluded that time gap plays a role in compensation for the room factor but not the loudspeaker factor. However, it is hypothesised that, based on the tends seen in this experiment, a wider rating scale may result in a similar effect for the loudspeaker factor.

**Effect of instructions**

The effect of instructions on compensation was almost non-significant for the loudspeaker factor (loudspeaker-by-condition-by-instruction interaction: $F = 3.138, p = .045$ partial eta squared $= .022$). The effect of instructions was significant for the room factor (room-by-condition-by-instructions interaction: $F = 7.549, p = .001$ partial eta squared $= .050$). However, for both factors compensation reduced because differences became smaller when compared directly with instructions rather than larger when stimuli were compared indirectly (and this cannot be explained by a corresponding increase in the range of the indirect factor, as was suggested for the time gap analysis). Therefore, the results are not suggestive of instructions causing real-world compensation. The small but significant reduction in variation when stimuli are rated directly should be explained. This may be explained by the listeners expecting to hear large differences across the test that did not materialise, and making space on the
rating scale for these.

### 3.18 Experiment 3: Summary and discussion

Experiments 1 and 2 and Olive et al.’s (1995) experiment show real-world compensation for loudspeakers and rooms when listening over a longer time course (i.e. when comparing channels indirectly). It was hypothesised that in particular longer time gaps are likely to play a role in this effect but non-time-gap sensitive mechanisms and task factors such as instructions may contribute to this effect. The data provided by this experiment showed significant compensation for both loudspeakers and rooms. Therefore, further evidence of compensation for both these channels in real-world listening is provided. An aim of this experiment was to determine whether bias caused by the task instructions is a cause of compensation. Specifically, it was hypothesised that the task set-up and the task instructions (rating the stimuli on a page by page basis) may have encouraged listeners to intentionally ignore differences across pages. A test of this bias was conducted by asking listeners to rate stimuli globally with regard to all stimuli across the test and any reduction in compensation was measured. Results showed that compensation did not reduce with instructions to make global ratings compared to no specific instructions. Therefore, an intentional recalibration of rating scales with each new page does not appear to behind the compensation seen in previous experiments. The compensation seen here appears to be a genuine listening phenomenon rather than an artefact of the task set-up and instructions. It appears that Olive et al. used similar instructions and so this result shows that this sort of response bias was probably not behind their results.

In the experiment the instructions to make global ratings aimed to encourage equal attention to the differences occurring across the task as well as on a single experimental page. Therefore, it might also be concluded that differences in the level of attention directed to stimuli when directly and indirectly compared do not explain compensation. The lack of effect of instructions suggests that listeners do not have control over observing differences in the indirect factor. They cannot do this even when asked to do so and it is not a question of lack of attention, which can be controlled and directed to
the task. However, attention may still be unequal between the two conditions because of a number of factors: listeners had control of changes in the direct comparison condition and changes occurred regularly for this factor but not for the indirect factor. Therefore, attention to differences between the direct factor is still likely to be increased compared to the indirect factor despite instructions to rate globally. Attention as an explanation for results could be researched further.

This experiment also tested whether compensation seen in the previous experiments was caused by time gaps between stimuli. The 0 minute time gap indirect comparison condition (where there was no silent gap between indirectly compared stimuli) was compared to the 10 minute time gap condition, and any reduction in compensation was measured. Time gaps between stimuli presentations were partly responsible for the compensation. Compensation reduced with the removal of time gaps for the room factor and this reduction is approaching ‘moderately large’, according to rating scale values. Importantly, the reduction was in the expected direction: removing breaks caused larger perceived differences in rooms with indirect comparisons. However, there is still a large amount of compensation left to explain. Non-time-gap sensitive mechanisms must cause the remaining compensation. The exact mechanisms behind the time-gap sensitive and non-time-gap sensitive components of compensation seen in this test are not known and should be investigated in further work.

For the loudspeaker factor the difference between direct and indirect comparisons also reduced with time gap but this reduction occurred mainly because of a decrease in perceived variation in loudspeakers when compared directly, rather than an increase in perceived variation with indirect comparisons. This is not expected by the compensation hypothesis. The failure to find a release from compensation with indirect comparisons (differences were still small with indirect comparisons) may have been due to the fact that any release was prevented from being displayed in ratings because there was not sufficient room on the rating scale due to the large differences between rooms. The ratings for the loudspeaker and room factor are in theory independent but the fact that the rating scale is limited means that the more range taken up by one factor, the more space on the scale is restricted for the other factor. Therefore, it should be concluded that removing the time gap does not appear to explain compensation for the loudspeaker factor, but the possibility of ceiling effects needs to be examined.
3.18 EXPERIMENT 3: SUMMARY AND DISCUSSION

It is possible that the same effect would be seen for loudspeakers if ceiling effects are prevented. However, it is not necessary to test this as part of this thesis as there is sufficient evidence of a time gap sensitive mechanism of compensation in this experiment.

If the results show genuine asymmetry between the mechanisms of compensation for loudspeaker and rooms, this may be because of differences in the way that the hearing system deals with these channels. In this research rooms and loudspeakers are treated as generic transmission channels that colour a sound in the same way. However, the way colouration from rooms reaches the listener is different to the colouration from a loudspeaker. Room colouration arrives via reflections from different directions, with different times of arrival and the colouration from each is different. Colouration from the loudspeaker is imparted onto the direct sound (and therefore all the reflections) in an equal manner. It is likely that different methods of compensation exist for the two types of channel colouration. For example binaural hearing may be used to reduce the effects of room reflections, or time of arrival cues may be used, which cannot be used for removing loudspeaker colouration. The current results may provide evidence of a difference in mechanisms of real-world compensation for these factors. Alternatively, the difference may be caused by the differences in the initial effect size. Ideally future experiments would aim to measure compensation for factors that have the same magnitude of variation when compared directly, so that the potential for compensation is equal for both factors. However, in practice it might to be difficult to achieve this. It is also noted that removing the break phase did not make the time between stimulus onsets equal between the directly and indirectly compared stimuli. Further, because of the restricted order of comparisons in the indirect condition some comparison between stimuli with other stimuli heard between them occurred (i.e. there would have been interference as well as time gap). This may have resulted in reduced sensitivity to timbral differences. A statistical comparison of differences in stimuli that were presented side-by-side versus non-side-by-side when indirectly compared was conducted to determine whether larger differences are seen when stimuli can be compared side-by-side but a lack of statistical power prevented conclusive analysis.

The results in this experiment are contrary to those of Bech (1994), who showed that stable perception of listening room preference occurs with a 2 month listening gap. Bech
(1994) concluded that the expert listeners used in his test were able to maintain long-term representations of room timbre. However, expert listeners were also used in the current test and this was not found. Further, a small tendency towards a contraction of ratings with time can be seen in Bech’s results. However, Bech’s finding highlights the fact that some form of long-term memory for timbre exists. Listeners are able to label loudspeakers and rooms as good and bad consistently over time. Total compensation does not appear to occur. It is likely that this memory involves the storage of timbre in a categorical form and is different from the memory involved in observing smaller timbral differences such as those being compared between loudspeakers in this test.

### 3.19 Experiment 3: Conclusion

The primary aim of this experiment was to determine the mechanisms of compensation. Further evidence for compensation was seen as loudspeaker and room differences were smaller with indirect compared to direct comparisons. It was shown that instructions to make global ratings did not result in decreased compensation. The fact that compensation still occurs when instructions are given to observe differences due to the changes in the indirect factor shows that listeners cannot perceive large differences between the indirectly compared stimuli even when instructed to pay attention to these. A bias towards rating directly compared differences as larger because of the task instructions/task set-up, can be ruled out. There are no other experimental factors that appear to cause compensation so it can be concluded that compensation is a genuine listening phenomenon. There was a small but significant effect of instructions which showed that listeners tended to be slightly more conservative in their ratings when comparing stimuli directly with instructions. This may show that, expecting to hear differences across the test, listeners saved space on the rating scale to reflect changes across the test.

The effect of comparison method was partly explained by the gaps between stimulus presentations with indirect comparisons, suggesting that a time-gap sensitive type of compensation mechanism is be involved in the compensation seen in previous experiments. This mechanism may be timbral memory but is more likely to be a
specific time-gap sensitive compensation mechanism. This time-gap sensitive effect was only shown for the room factor but was approaching moderately-large for this factor. The lack of effect of time gap for the loudspeaker factor may have been due to the fact that room differences were perceived to be large in this test when rated directly and indirectly. It may be that large differences in one factor suppress increases in the other factor that may have otherwise occurred. This may show a compensation that is in part due to stimuli needing to share a limited perceptual scale or might show a limitation of the test methodology due to the limited range of the rating scale in use in these tests. This effect could be investigated further as an explanation for compensation. Future tests using the real-world paradigm should aim to confirm that no time-gap sensitive compensation occurs for the loudspeaker factor when the possibility of ceiling effects are reduced or removed. It remains possible that the lack of effect for the loudspeaker factor shows that a different mechanism of compensation occurs for loudspeakers and rooms or that the size of the initial magnitude of the channel effect affects the overall magnitude of compensation that occurs and the contribution of time-gap sensitive component of compensation. However, at present it is concluded that time-gap sensitive compensation would likely be shown for this factor as well as the room factor if the possibility of ceiling effects was removed. It can be concluded that a time-gap sensitive mechanism contributes to compensation for the spectral effects of loudspeakers and rooms and further experiments are necessary to determine the exact mechanisms behind this time-gap sensitive component of compensation. It can also be concluded that, because a large amount of compensation remains after the time gap is eliminated, a non-time-gap sensitive mechanism of compensation contributes to compensation and this could also be investigated in future work.

3.20 Chapter summary and discussion

In Chapter 2 an experiment by Olive et al. (1995) that showed compensation for channel colouration in a real-world listening scenario was discussed. It was noted that further experiments were needed to confirm this real-world compensation using Olive et al.’s (1995) paradigm and to measure its extent. This replication was necessary because real-world compensation had only been properly shown in this single previous test. Further,
it was noted that once compensation was confirmed, this paradigm could be used to test the mechanisms behind compensation. In the current chapter three experiments aiming to replicate and expand the research of Olive et al. (1995) were conducted. Experiment 1 aimed to confirm that real-world compensation occurs and examine its extent. Compensation was only significant at the p<.10 level. Therefore Experiment 2 was conducted under conditions more conducive to compensation. Compensation was large to moderate. Experiment 3 further investigated compensation mechanisms. Chapter questions were presented which were to be answered via the work presented. These questions aimed to guide the work in this chapter towards providing information necessary to answer Research Questions 1a and 1b. The findings of this chapter are therefore summarised and discussed in the course of answering the chapter questions below. The extent to which Research Questions 1a and 1b can be answered from this work is discussed in brief in the conclusion to this section and in the conclusion to this thesis (see Chapter 7).

3.20.1 Chapter questions

The answers to the chapter questions are addressed:

Experiments 1 and 2 answered Question 3.1. ‘Can compensation for loudspeakers and rooms seen in the real-world experiments described in chapter 1 be elicited again with new stimuli?’ Experiment 1 showed evidence of real-world compensation for the loudspeaker factor but this was only significant at p<.10 not the stricter p<.05. Compensation for the room factor could not be measured because of floor effects. The variation in rooms was not perceived to be large enough to be significant so could not reduce further with compensation. A trend towards compensation for the room factor was seen, however. It was concluded that significant compensation for both factors at p<.05 should be obtained to confirm that real-world compensation occurs using Olive et al’s paradigm. In order to attempt to make timbral differences more audible and prevent these floor effects Experiment 2 used a speech programme item. Listening lengths and time gaps between stimuli were also increased in the indirect comparison condition to attempt to increase compensation. This resulted in highly significant differences in both loudspeakers and rooms when rating directly and highly
significant compensation for both the loudspeaker and room factors. Compensation in this experiment was ‘large’ (29 preference points) to ‘moderate’ (10 preference points) according to rating scale values. Experiment 3 also showed compensation. Overall, there is now evidence of compensation occurring with speech and non-speech programme items, both replicating and increasing the validity of Olive et al. (1995) et al’s results to another programme item. However, it was noted that it might be informative to conduct tests which make direct comparisons of compensation with both programme items as compensation may vary in its extent between programme items. Such tests could be conducted as part of work following on from this thesis. It can be concluded that real-world compensation is confirmed and its extent measured.

Question 3.3 asked ‘what type of perceptual mechanisms might be behind this compensation?’ It is evident that it is possible to manipulate the real-world paradigm to test for mechanisms of compensation as was done in Experiment 3. The results of previous experiments suggested that two factors were particularly important in causing this compensation: 1) the length of continuous listening to the channel, possibly because of fatigue in response to continuously presented spectral energy (as seen in loudness adaptation (Hood 1950)) and; 2) the time between comparisons, possibly due to timbral memory or other time-gap sensitive compensation mechanisms. In Experiment 2 a preliminary test of the role of longer listening could be carried out (the experiment did not aim to examine this but it was possible to examine this to a limited extent). Increased compensation did not occur between approximately 25 s to 4 minutes of listening suggesting that longer listening does not cause compensation. However, if compensation is caused by a process that is similar to loudness adaptation it would be expected to occur within a relatively short time course (e.g. between 0 and 25 s) because adaptation-based mechanisms occur more readily within this time period. Therefore, this result does not rule out longer continuous listening as a cause of real-world compensation. Future work could investigate the role of longer listening before 25 s further. However, this was not investigated further here. Instead, Experiment 3 aimed to investigate Olive et al.’s preferred explanation and determine whether a type of mechanism that is sensitive to time gaps between comparisons causes compensation. To measure this a condition was run with the time gap between indirect comparisons removed. Compensation was moderately and significantly larger
with 10-minute gaps for the room factor compared to no time gaps but the time gap did not appear to have an effect on compensation for the loudspeaker factor. It is possible that the lack of effect for this factor was due to ceiling effects and that time gaps would also explain compensation for this factor if these were not present. Therefore, the role of time gap in compensation is confirmed for rooms but results are inconclusive for loudspeakers. Overall, enough evidence exists to conclude that a time-gap sensitive component of real-world compensation exists and this must be caused by time-gap sensitive mechanisms. Compensation remains after the time gap is removed. Mechanisms that are not affected by the time between stimuli (non-time gap sensitive mechanisms) must be behind this component of compensation. It was concluded that little is currently known about what specific mechanisms might underlie the time-gap and non-time gap components of real-world compensation and further work will be conducted as part of this thesis to investigate the causes of compensation.

In order to answer Chapter Question 3.2: ‘Is this compensation an artefact of the experimental process or is it likely to be a genuine listening phenomenon?’, it was necessary to confirm that compensation is not due to task factors—namely the instructions given to participants. Listeners were either instructed to make global ratings and mark any differences heard across experimental pages in their ratings or were given the same instructions as used in previous tests. There was a significant effect of instructions but instructions did not appear to explain reduced loudspeaker and room differences when rated indirectly (i.e. compensation). Instructions caused a slight decrease in perceived channel differences when rated directly. It appears listeners were making space on the rating scale to note differences across the test that they expected but that they did not eventually hear due to compensation. The lack of ability to observe differences between indirectly compared stimuli even when instructed to pay attention to these so also indicates that there is no role for attention in compensation. However, attention still remains unequal between direct and indirect comparisons so attention as a cause of compensation could be investigated further through future testing.
3.21 Chapter conclusion

The experiments presented replicated and confirmed real-world compensation for loudspeaker and rooms. Experiment 1 shows nearly significant compensation and Experiment 2 shows significant and moderate to large compensation. Therefore Research Question 1a has been answered. Research Question 1b was also partially answered. The instructions given to the listener do not appear to be the cause of the compensation effects so it can be said that the apparent compensation is not an artefact of the experimental process but a genuine phenomenon of real-world listening. Time gaps between comparisons play a role in compensation suggesting a type of mechanism that is time-gap sensitive causes compensation. However, compensation remained even after time gaps were removed so there is also role for mechanisms that are non-time-gap sensitive (possibly a mechanism that is sensitive to the length of listening under 25 s). The specific psychological and physiological mechanisms behind the time-gap and non-time-gap sensitive components of compensation are unknown and require further research. Currently little is known about mechanisms that might underpin this real-world compensation. A couple of speculative mechanisms have been suggested (memory and timbral fatigue) but these are not well-known compensation mechanisms. There may be specific compensation mechanisms that exist that can explain real-world compensation. Further work to determine the mechanisms behind time-gap and non-time-gap sensitive components of real-world compensation are discussed in the next chapter.
Chapter 4

Testing the mechanisms of compensation for loudspeakers and rooms

In the previous chapter compensation for loudspeakers and rooms was confirmed and the mechanisms behind this were investigated. Experiment 3 showed a role for a type of mechanism that is dependent on the gap between stimulus comparisons and a type of mechanism not dependent on these time gaps. It was acknowledged that a better understanding of specific mechanisms of compensation is needed in order to determine the mechanisms of real-world compensation further. The work described in the rest of this thesis will contribute to determining specific mechanisms that have the potential to contribute to real-world compensation. A research process is set out in this chapter to do this. The research in the remaining chapters therefore aims to answer Research Question 1b:

‘By what means could compensation affect the perception of loudspeaker and rooms?—What are the potential mechanisms of real world compensation for loudspeaker and rooms?’.

As noted in Chapter 1, Question 1c is also of interest:

Research Question 1c: ‘What are the actual mechanisms of real-world compensation
for loudspeaker and rooms?’

Question 1c will not be answered directly by this thesis but as well as a research process to determine potential mechanisms of real-world compensation, a process to determine the actual mechanisms of compensation is presented in this chapter. This is done in order to give the reader an understanding of how the work conducted will ultimately help to determine the actual mechanisms of compensation in real-world listening. As in previous chapters, chapter questions are set out to guide the work in this chapter towards providing answers to the research questions:

**Question 4.1:** What research is necessary to further determine potential mechanisms of compensation for loudspeakers and rooms when listening to speech and non-speech in the real world?

**Question 4.2:** To what extent will the research in this thesis contribute to determining this?

**Question 4.3:** How will actual compensation mechanisms be determined?

The current chapter describes a two-step research process to determine the potential mechanisms of real-world compensation (Step 1) and the actual mechanisms of compensation (Step 2). Section 4.1 describes this two-step research process and the rationale for a two-step process. Section 4.2 gives a skeleton structure for this process and describes which chapters in the remainder of thesis fulfil which parts of the process. Section 4.3 provides a summary and discussion of the work in this chapter and provides answers to the chapter questions. The the extent to which the main research questions can answered by the work in this chapter is also discussed here and in Chapter 7.

### 4.1 A two-step selection and elimination process

Types of mechanisms behind real-world compensation have been established in Experiment 3 (time-gap sensitive and non-time-gap sensitive), but it is not clear what the specific mechanisms behind real-world compensation are and whether there are known compensation mechanisms, that have been investigated in other auditory
4.1 A TWO-STEP SELECTION AND ELIMINATION PROCESS

research domains, that may explain this real-world compensation. The time-gap sensitive component of compensation could be due to a process that enhances differences between sounds when heard close in time but not when heard further apart. Memory could play a role in this, but specific compensation mechanisms may also be behind this. Another explanation for the real-world compensation seen in Olive et al.’s (1995) experiment and the experiments in Chapter 3 could be the longer listening periods. A fatiguing effect, whereby the listener becomes less sensitive to timbre with time spent listening was suggested. Further research into known specific compensation mechanisms is necessary to determine whether any of these are a cause of the compensation seen in real-world tests. A search for compensation mechanisms should be conducted to determine whether any of these can explain real-world compensation.

A number of previous psychoacoustic laboratory studies describe mechanisms of compensation for channel colouration that may explain real-world compensation (e.g Watkins (1991), Summerfield et al. (1987)—these and other studies will be described in Chapter 5). However, these psychoacoustic studies tend to involve testing compensation mechanisms over short listening time courses (millisecond and seconds) and in situations that are unlike real-world listening (e.g. listeners hear isolated sounds rather than multiple factors at once). Therefore, little is known about how these mechanisms contribute to the real-world compensation seen in the experiments in Chapter 3 and in Olive et al.’s (1995) experiment. The mechanisms seen in these psychoacoustic laboratory studies might be involved in real-world compensation, but due to the shorter time courses under which they are usually tested and other differences between real-world and laboratory listening, they may not contribute, or only contribute a small amount to the overall compensation seen in real-world tests. There may be other mechanisms that occur in real-world listening but not in laboratory scenarios which are more important to real-world compensation. Research which aims to determine the actual mechanisms of real-world compensation firstly involves a process of selection of potential compensation mechanisms from the psychoacoustic laboratory studies. A literature review of psychoacoustic studies which appear to show mechanisms involved in compensation for the spectral effects of the channel should be undertaken and a list of mechanisms that might explain real-world compensation should
be drawn up. Using this list of potential real-world compensation mechanisms, each mechanism can be tested for its actual role in real-world compensation via a process of elimination. This will determine whether each potential mechanism contributes to the real-world compensation and to what extent. Any remaining compensation after this process of elimination is complete will show that other mechanisms of compensation must be behind real-world compensation (i.e. mechanisms of compensation that have not previously been found in psychoacoustic laboratory research and may only occur in real-world listening).

This research process is therefore divided into two steps. Step 1 involves a selection of potential mechanisms of real-world compensation via a literature review of compensation mechanisms that have been established using psychoacoustic laboratory studies, and also where necessary via additional tests conducted as part of this thesis. Step 2 involves determining the actual role of each of these mechanisms in real-world listening. The steps are described in more detail below:

**Step 1: Selection—Determining potential mechanisms of compensation**

A list of mechanisms shown to cause compensation for channel colouration in psychoacoustic laboratory tests is compiled through a literature review. As real-world compensation was shown to be caused by time-gap sensitive and non-time-gap sensitive mechanisms, this list will be divided into those mechanisms that are time-gap sensitive and those that are not. For each mechanism on the list, where there is sufficient data to answer Questions A-C (below) in the affirmative, then it is possible to conclude that this mechanism is a *potential* mechanism of real-world compensation for loudspeakers and rooms. This mechanism can be added to a list of potential mechanisms of real-world compensation.

**Step 1 questions:**

A) For any compensation mechanism, does the mechanism produce a measurable effect that is indicative of compensation and compatible with providing the real-world compensation seen in Olive et al.’s (1995) experiment and in the experiments described in Chapter 3?
4.1 A TWO-STEP SELECTION AND ELIMINATION PROCESS

B) Is there one or more distinguishing feature of this mechanism (e.g. is it crossaural or does it require a long onset)? This is required to:

- confirm that the mechanism is a unique and distinct mechanism and not a manifestation of another compensation mechanism and
- separate this compensation mechanism from other compensation mechanisms in tests of elimination in Step 2.

C) Is the time course of the mechanism such that it is compatible with being an explanation of the real-world compensation seen in Olive et al.’s (1995) experiment and in the experiments described in Chapter 3?

Where Step 1 Questions A-C can be answered for a number of mechanisms and a list of potential real-world compensation mechanisms can be complied, Research Question 1b will be answered. It has been suggested that compensation mechanisms may be different when listening to speech compared to other sounds (e.g. Liberman et al. (1967), see Chapter 5 for a discussion). For each potential mechanism Questions A-C must be answered for speech and non-speech listening so that potential mechanisms of real-world compensation for loudspeakers and rooms when listening to all sounds, including the music used by Olive et al., can be established.

In some cases all the information necessary to answer Questions A-C for the mechanism in question will be provided by the previous psychoacoustic research described in the literature review. However, it is expected that for some mechanisms flagged up by the literature there will be insufficient data to answer Questions A-C in the affirmative for speech and non-speech listening. In this case additional laboratory experiments will be conducted as part of this thesis to provide this information (see Chapter 6 for details of these tests). Conducting these experiments to answer Questions A-C will involve using the psychoacoustic laboratory paradigm. Therefore, as well as providing answers to Questions A-C, the tests will fill gaps in the psychoacoustic literature and provide a useful addition to the existing psychoacoustic data on compensation. Further, the information on compensation from these additional laboratory studies can be compared with previous research more directly because both use the laboratory paradigm. Therefore, checks can be made by comparing the results of this research
4.1 A TWO-STEP SELECTION AND ELIMINATION PROCESS

with previous studies.

When this data is obtained and Step 1 Questions A-C answered it will be possible to determine which mechanisms can be added to the list of mechanisms with the potential to cause compensation in real-world listening. The process in Step 1 is mainly one of a selection of potential mechanisms of real-world compensation but equally it may be possible to eliminate some of the mechanisms of real-world compensation at this stage (i.e. those that do not fulfil Questions A-C can be eliminated as possible explanations of real-world compensation). Elimination at this stage offers the benefit that it will not be necessary to run a test using the more complex real-world paradigm to examine the role of this mechanism in real-world compensation. Once the list of potential mechanisms is complete and Research Question 1b answered, Research Question 1c can be addressed and each can mechanism can then be tested for their actual role in real-world compensation in Step 2.

Step 2: Elimination—Determining actual mechanisms of compensation

Step 2 describes a research process for determining the actual mechanisms of compensation. This step will not be fulfilled as part of this thesis but a brief outline of the research process is given here and described further in Chapter 7. It will be seen that all the potential mechanisms of real-world compensation have only been tested in psychoacoustic laboratory experiments, which may not use conditions that are similar to real-world listening (such as longer listening time courses). Mechanisms shown to be potential mechanisms of real-world compensation via information obtained from laboratory listening may prove to be relatively unimportant means of compensation in real-world listening. This step therefore involves eliminating each of the potential mechanisms revealed in Step 1 using the real-world listening paradigm to determine which have a significant role in real-world listening. This process of elimination follows the process that was started in Experiment 3 in Chapter 3 where time gaps and instructions were tested for their role in real-world compensation by eliminating these causes (removing the time gaps and eliminating the possibility for bias by altering the instructions) and measuring extent to which compensation reduced. The unique features of each potential mechanism identified in Step 1 (Question B) will be used to single out mechanisms for this process of elimination and the relative role of each
mechanism in real-world compensation will be examined. This process continues until all of the potential compensation mechanisms have been tested and the extent to which they contribute to real-world compensation confirmed.

In conducting Step 2 questions D-F will be posed:

D) Does the potential real-world compensation mechanism contribute to the compensation that is seen using the real-world paradigm as well as the compensation seen in laboratory tests?

E) What is the extent of the contribution of each of the potential mechanisms to real-world compensation?

F) Has real-world compensation been fully explained by the contribution of the potential mechanisms drawn from laboratory research or do mechanisms not revealed in laboratory listening play a role in compensation when listening over longer time courses.

4.2 Research process skeleton plan

Below is a skeleton outline of the whole research process. This is divided into steps 1 and 2 and sub-steps which are presented in the order in which they should be conducted. The whole process is described but only part of this process will be completed in this thesis. As described above, potential mechanisms of real-world compensation can be divided into those that explain compensation caused by the time gap between sounds and those that explain the compensation that remains after this has been removed. It should be possible to investigate both of these time-gap and non-time-gap components of compensation separately (though in practice it may be revealed that there is some overlap with mechanisms contributing to both components of real-world compensation). To divide the work appropriately, it is decided that rather than completing Step 1, then Step 2, both steps should be conducted for the time-gap sensitive mechanisms first and this will be followed by conducting both steps for the non-time-gap sensitive mechanisms. Time-gap sensitive mechanisms are researched first because it will be seen that more advancement has been made by previous work
on compensation mechanisms that may be considered time-gap sensitive. Therefore, we are closer to determining time-gap sensitive sources of real-world compensation for loudspeaker and rooms. Less is known about mechanisms that might be non-time-gap sensitive and more work is needed to determine the potential and actual mechanisms behind the non-time-gap sensitive component of real-world compensation.

It is preferable to complete our knowledge of the time-gap sensitive component of real-world compensation before investigating the non-time-gap sensitive mechanisms. Therefore, to limit the scope of this thesis only the time-gap sensitive component of compensation is investigated. However, the process of determining the role of the both time-gap and non-time-gap sensitive mechanisms is outlined in the plan below and in the course of the literature review a limited discussion of non-time-gap sensitive mechanisms will be given to give the reader an understanding of all mechanisms that might be involved in real-world compensation.

As well as limiting the work in this thesis to investigating the time-gap sensitive component of compensation, only Step 1 (the search for potential mechanisms) is conducted. The investigation of actual time-gap sensitive mechanisms of compensation (Step 2) will be conducted in future work and is discussed again in Chapter 7.

**Skeleton plan**

The Step 1 and 2 process is conducted separately for a) the time-gap sensitive mechanisms and b) non-time-gap sensitive mechanisms. All of Step 1a is conducted in this thesis. Step 2a, Step 1b and Step 2b will not be conducted by the work in this thesis but experiments that work towards completing these steps will be discussed in Chapter 7.

**Conducted as part of this thesis:**

**Step 1a)** Determining potential time-gap sensitive mechanisms of real-world compensation via:

- Step 1a(i), a literature review investigating time-gap sensitive compensation mechanisms revealed in previous laboratory tests. This will be conducted for speech and non-speech listening to determine potential mechanisms of real-world
compensation when listening to either programme type. See Chapter 5.

- Step 1a(ii), additional laboratory tests to further determine potential time-gap sensitive compensation mechanisms. The information uncovered in the literature review may be sufficient to answer Step 1 Questions A-C for speech and non-speech in the affirmative, and determine a list of potential time-gap sensitive mechanisms of compensation. Where the information is not adequate to establish potential mechanisms of real-world compensation with speech and non-speech, additional experiments will be conducted to provide this information. This will be done for speech and non-speech listening, as required. See Chapter 6.

To be conducted as part of future work

The following steps will not be conducted as part of this thesis. Some of the work contributing to fulfilling these steps is discussed further in Chapter 7.

Step 2a) Determine the actual role of the time-gap sensitive mechanisms in real-world compensation: Each of the potential time-gap sensitive mechanisms of compensation will be tested for their role in real-world compensation using the real-world paradigm and tests of elimination. This will be conducted for speech and non-speech listening.

Step 1b) Determining potential non-time-gap sensitive mechanisms via:

- Step 1b(i), a literature review investigating non-time-gap sensitive compensation mechanisms revealed in previous laboratory tests. This will be conducted for speech and non-speech listening to determine potential mechanisms of real-world compensation when listening to either programme type.

- Step 1b(ii), additional laboratory tests to further determine potential non-time-gap sensitive compensation mechanisms. The information uncovered in the literature review may be sufficient to answer Step 1 Questions A-C in the affirmative for speech and non-speech, and determine a list of potential non-time-gap sensitive mechanisms of compensation. Where the information is not adequate to establish potential mechanisms of real-world compensation with speech and non-speech, additional experiments will be conducted to provide this information. This will be done for speech and non-speech listening, as required.
Step 2b) Determine the actual role of non-time-gap sensitive mechanisms in real-world compensation: Each of the potential non-time-gap sensitive mechanisms of compensation will be tested for their role in real-world compensation using the real-world paradigm and tests of elimination. This will be conducted for speech and non-speech listening.

4.3 Chapter summary and discussion

In the previous chapter it was concluded that time-gap sensitive mechanisms of compensation are involved in real-world compensation and non-time-gap sensitive mechanisms also contribute. The current chapter described the aim of the remaining chapters in this thesis—to contribute to determining the mechanisms of real-world compensation for loudspeaker and room colouration when listening to speech and non-speech. A research process to do this was described. This involves 1) determining potential mechanisms of real-world compensation then 2) actual mechanisms of real-world compensation.

It was acknowledged that this two-step process is necessary because potential mechanisms of compensation must be searched for so that these can be eliminated to determine actual mechanisms of compensation. This search is necessary because little is currently known about which mechanisms may contribute to real-world compensation, therefore it is not possible to eliminate these explanations directly by continuing to use the real-world paradigm (as was done in Experiment 3). Potential mechanisms of compensation can be determined by looking at compensation mechanisms reported in previous psychoacoustic literature. It was noted that compensation mechanisms that may be reported in the psychoacoustic literature will have only been tested in psychoacoustic laboratory experiments, which tend to use short time courses, and other situations that are unlike real-world listening. The extent to which any such mechanisms provide a significant or important contribution to real-world compensation is therefore not clear. Of the mechanisms that are reviewed a list of those that may be potential mechanisms will be set out. Those mechanisms that A) produce compensation effects that show they have a nature suitable to cause the real-world compensation, B)
have distinguishing features that show that they are unique mechanisms and not a manifestation of another mechanism and C) have a time-course suitable to cause the real-world compensation with speech and non-speech listening, may be regarded as potential mechanisms of compensation for loudspeakers and rooms. These specific potential mechanisms can then be tested further for their actual role in compensation in real-world listening (Step 2). This will be done using their unique features, as established by the literature review, to eliminate each in tests using the real-world listening paradigm. Such tests of elimination will determine the extent of role of each potential mechanism to real-world compensation.

It is expected that the literature review will not provide sufficient information to answer Questions A-C for all mechanisms flagged up by the literature review for both speech and non-speech listening. In this case additional laboratory tests will be conducted to obtain this information. It is also expected that the mechanisms seen in the previous psychoacoustic laboratory tests may only explain part of real-world compensation; there may be other mechanisms that occur in real-world listening that are not revealed in those tests. Any compensation not explained by the potential mechanisms listed may be due to mechanisms that have not been revealed in psychoacoustic laboratory studies (perhaps because of the shorter time courses researched) but occur in real-world listening. Further, the remainder of this thesis focuses on determining the mechanisms behind the component of real-world compensation that is time-gap sensitive. In Chapter 3 it was noted that a number of specific time-gap sensitive compensation mechanisms may be found in the course of the literature review that might explain the time-gap sensitive component of compensation and the research process should focus on testing the role of any of these mechanisms in the real-world compensation before other more speculative mechanisms, such as timbral memory and timbral fatigue, are considered.

4.3.1 Chapter questions

The chapter questions can be answered:

Question 4.1 asked: 'What research is necessary to further determine potential
mechanisms of compensation for loudspeakers and rooms when listening to speech and non-speech in the real world?’. A literature review is necessary to identify potential mechanisms of compensation so that the contribution of known compensation mechanisms to real-world compensation can be tested using the real-world paradigm. Mechanisms from the literature review will be regarded potential mechanisms of real-world compensation if Step 1 Questions A-C can be answered for speech and non-speech listening. Step 1 Questions A-C require that the mechanisms are unique and have a nature and time course that suggests that they may explain the reduced sensitivity to the channel seen with indirect comparisons in real-world experiments.

**Question 4.2** asked about the extent to which the research in this thesis contributes to determining mechanisms of real-world compensation. The current work will aim to determine a list of potential mechanisms of real-world compensation via a literature review and experiments. However, it will only investigate mechanisms behind the time-gap sensitive component of real-world compensation, rather than the component of compensation caused by non-time-gap sensitive mechanisms. It will only determine potential rather than actual mechanisms and it will only investigate the extent to which the psychoacoustic laboratory mechanisms play a role in real-world compensation, not the extent of the contribution of any mechanisms that act in the real world but have not been tested in psychoacoustic laboratory tests. In summary, only Step 1a is conducted in this thesis, leaving Step 1b and Step 2a and 2b for future work.

**Question 4.3** asked: ‘How will actual compensation mechanisms be determined?’. Ultimately, the role that the mechanisms actually play in real-world compensation should be tested. Potential mechanisms can be tested for their actual role in compensation via a process of using their unique features to exclude their contribution to real-world compensation (Step 2). This process of elimination should continue until the contribution of all the potential mechanisms is ascertained. Any remaining compensation after the exclusion of all potential mechanisms examined here shows that other mechanisms contribute to real-world compensation. These may be mechanisms that work over a longer time course than used in the previous psychoacoustic laboratory work or mechanisms that have not been apparent in laboratory studies for other reasons. These tests may be conducted as part of future research but are outside of the scope of this thesis.
4.4 Chapter conclusion

The research questions were not directly addressed in this chapter but a research process was set out to provide further answers to Research Questions 1b and 1c. This research process aims to determine the mechanisms behind the real-world compensation for loudspeakers and rooms that was seen Olive et al.’s (1995) study and the Experiments described in Chapter 3. The remainder of this thesis will fulfil part of this research process and will determine potential mechanisms behind the time-gap sensitive component of real-world compensation (as shown in Experiment 3). A literature review will be conducted to determine mechanisms that compensate for the spectral effects of the channel as revealed by psychoacoustic studies. Of these mechanisms those that are unique and have the nature and time course appropriate to explain real-world compensation for the spectral effects of the channel may be designated potential mechanisms of real-world compensation and Research Question 1b will be answered. However, it is likely that insufficient information to determine whether the mechanisms are potential mechanisms of real-world compensation will be obtained from previous research. Additional laboratory tests may needed to obtain this information. These will be conducted as part of this thesis if necessary. Once a list of potential mechanisms is compiled future work is outlined which will contribute to determining the extent of the contribution of these mechanisms to real-world compensation using the real-world test paradigm.
Chapter 5

Potential mechanisms of real-world compensation: speech and non-speech

In the previous chapter a research process to determine potential and actual mechanisms of real-world compensation was set out. The next step in this research process is to determine potential mechanisms of real-world compensation via a literature review. This review is described in the current chapter. Specifically, in order to limit the scope of the work in this thesis, this review will focus on identifying potential time-gap sensitive mechanisms of compensation, which can explain the time-gap sensitive component of real-world compensation which was shown to occur in Experiment 3 in Chapter 3. As was explained in Section 4.2, the investigation of mechanisms behind the non-time-gap sensitive component of real-world compensation is reserved for future work. The current work therefore aims to fulfil Step 1a(i) of the research process set out in Chapter 4. However, in the course of identifying such mechanisms non-time-gap sensitive mechanisms are also discussed. If after completing the review, a number of mechanisms that compensate for channel colouration can be identified, these are shown to be time-gap sensitive and Step 1 Questions A-C (see Section 4.1 and below) can be answered for these mechanisms for speech and non-speech listening, then potential mechanisms of real-world compensation will have been identified and Research Question 1b answered. However, where it is not possible
to confirm time-gap sensitivity or answer Questions A-C with information from the literature review, additional experiments will be conducted to obtain the necessary information (i.e. Step 1(a)ii will be conducted—see Chapter 6). Step 1 Questions A-C are as follows:

Step 1 Questions:

A) For any compensation mechanism, does the mechanism produce a measurable effect that is indicative of compensation and compatible with providing the real-world compensation seen in Olive et al.’s (1995) experiment and in the experiments described in Chapter 3?

B) Is there one or more distinguishing feature of this mechanism (e.g. it is crossaural or does it require a long onset)? This is required to:

- confirm that the mechanism is a unique and distinct mechanism and not a manifestation of another compensation mechanism and
- separate this compensation mechanism from other compensation mechanisms in tests of elimination in Step 2.

C) Is the time course of the mechanism such that it is compatible with being an explanation of the real-world compensation seen in Olive et al.’s (1995) experiment and in the experiments described in Chapter 3?

Literature on mechanisms that might contribute to the time-gap sensitive component of real-world compensation for the spectral effects of the channel comes from a variety of studies. These studies investigate different aspects of speech and non-speech perception. Broadly the studies concern:

- Compensation for the spectral effects of the vocal tract and the phonetic context;
- General mechanisms of compensation for the spectral effects of the channel in speech perception; and
- General mechanisms of compensation for the spectral effects of the channel in non-speech perception.
A review of the compensation mechanisms described in these different types of studies must be conducted. To structure the literature review towards identifying time-gap sensitive compensation mechanisms and providing answers to Questions A-C for each type of study, chapter questions are set out. These will be answered using the research arising from each of the above domains.

**Question 5.1:** a) To what extent does compensation for the spectral effects of the vocal tract occur in speech perception? b) What are the mechanisms behind this? c) What is the time course of this compensation? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

**Question 5.2:** a) To what extent is the phonetic context compensated for? b) What are the mechanisms behind this compensation? c) What is the time course? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

**Question 5.3:** a) To what extent is the effect of transmission channels other than the vocal tract (e.g. loudspeakers, rooms) compensated for in speech perception? b) What are the mechanisms behind this? c) What is the time course of effects? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

**Question 5.4:** a) To what extent does compensation for transmission channels occur when listening non-speech sounds? b) What mechanisms are behind this? c) What is the time course of effects? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

In light of the research coming from 3 domains, the literature review is divided into three sections and the chapter questions will be answered at the end of each. There is a further section to determine which of the mechanisms described throughout the review can be considered potential mechanisms of the time-gap sensitive component of real-world compensation (i.e. to determine whether Questions A-C and time gap sensitivity are confirmed for each mechanism) and there is a conclusion section. The sections are as follows:
5.1 LITERATURE REVIEW PART 1: COMPENSATION FOR THE VOCAL TRACT AND PHONETIC CONTEXT

1) **Compensation for the vocal tract (VT) and phonetic context in speech perception.** This section will primarily provide answers to Chapter Questions 5.1 and 5.2 (see Section 5.1).

2) **General compensation for the channel in speech perception,** which will primarily provide answers to Chapter Question 5.3 (see Section 5.2).

3) **General compensation for the channel in non-speech perception,** which will primarily provide answers to Chapter Question 5.4 (see Section 5.3).

4) **Establishing potential mechanisms of real-world compensation,** which will discuss the mechanisms revealed by the literature review, whether Questions A-C can be answered for any of these mechanisms and whether they are time-gap sensitive and therefore, whether they can be regarded as potential mechanisms of real-world compensation (see Section 5.4). Further work may be needed to confirm time-gap sensitivity and the answers to Questions A-C for each mechanism, and thereby answer Research Question 1b. This further work is discussed in Section 5.4 and some of this work is presented in Chapter 6.

5) **Chapter Conclusion.** This section concludes this chapter by summarising the findings of this chapter and the plan for future work (see Section 5.5).

5.1 Literature review part 1: Compensation for the vocal tract and phonetic context

There are few studies that aim to determine the physiological and psychological mechanisms behind compensation for transmission channels generally. Speech research provides most of the data on compensation for the spectral colouration caused by the channel. Specifically, research into compensation for the the spectral effects of a talker’s vocal tract on speech timbre and compensation for the spectral effects of the phonetic context (immediately adjacent speech made by the same talker), provides most of the data relating to this kind of channel compensation. This research is discussed with the aim of identifying mechanisms of compensation that might be involved compensation
for the spectral effects of loudspeakers and rooms in real-world listening when listening to music as well as speech.

In speech perception the vocal tract (VT) is the primary channel that modifies the spectrum of speech. The frequency response of the VT varies between people, due to differences in its size and shape. In Chapter 1 it was shown that the VT modifies the spectrum of speech produced through it to produce different speech sounds. The perception of phonemes was shown to occur when formant frequencies are produced by the talker. Because the spectrum of speech varies between different talkers because of VT size and shape limitations, the locations of formants are shifted between talkers. However, listeners appear to compensate for this VT colouration to maintain stable perceptions of speech sounds across talkers and produce timbral constancy. The immediately prior segment of speech, made by the same talker, also affects the spectrum of speech in a similar manner to VT variation across talkers. This effect of this prior phonetic context, appears to be compensated for in a similar way to VT variation.

The exact mechanisms behind this timbral constancy have been the subject of considerable research in speech perception. A number of potential mechanisms for producing this constancy will be discussed in this section including: the use of spectral cues that are relatively robust against VT variation, the role of non-spectral (temporal) cues in perception, the use of spectral cues within each phoneme that vary predictably across talkers to provide anchors for normalisation (compensation) and, compensation for VT differences via experience with the talkers spectral range from prior speech (via information outside of the phoneme). These mechanisms may also provide constancy when spectral colouration to speech is caused by other channels (such as loudspeakers), as well as constancy for VT colouration. Speech research provides most of the data on channel compensation but the extent to which compensation for colouration is the same for non-speech is mentioned in this section but examined specifically in Section 5.3

The first subsection of this section provides further information on the acoustical cues involved speech perception, which were introduced in Chapter 1 (see Subsection 5.1.1). The next subsection further discusses how the vocal tract can be considered a transmission channel which colours the sounds passing through it, and how stable
speech cues and intrinsic compensation mechanisms may exist to combat the effects of this colouration (see Subsection 5.1.2). This is followed by a discussion of extrinsic compensation mechanisms that may contribute to compensation for VT colouration in Subsection 5.1.3. In this Subsection the way in which the phonetic context also colours speech in a similar manner to the VT is described and extrinsic mechanisms for compensation for this source of colouration are put forward. Two further processes that may reduce spectral colouration caused by VTs and the phonetic context, Categorical Perception and Categorisation are described in Subsection 5.1.4. Subsection 5.1.5 provides a summary and discussion of the work in this section and answers the chapter questions. Subsection 5.1.6 concludes this section.

### 5.1.1 Phoneme perception and vocal tract colouration

In Chapter 1 the perception of vowel sounds was shown to be mainly determined by their timbre and in particular the location of formant frequencies (spectral peaks in the resonance of the vocal tract). These formant locations are cues to the recognition of most speech sounds but there are around 40-60 phonemes in the English language and many of these are allophones with more than one pronunciation (Ainsworth 1988). A number acoustical cues are thought to be relevant to the perception of sounds other than monophthongal vowels. These can broadly be divided into spectral (e.g. formants), temporal (e.g. voice onset times) or dynamic (changes in spectrum over time—e.g. formant transitions).

Fant (1973) defined the necessary timbral targets in speech perception as belonging to a number of broad categories: ‘It must be perceived whether or not the sound in question is vocalic/non-vocalic, consonantal/non-consonantal, compact/diffuse (related to the distance between F1/F2), grave/acute (related to position of F1/F2 average), flat/plain, nasal/oral, continuant/interrupted, strident/mellow’. Speech science has tried to determine the individual cues of speech more precisely. For example, the way that formants transition within phonemes has been shown to be particularly important to perception. For example, perception of the consonant /d/ is cued by formant transitions. Figure 5.1 (taken from Liberman et al. (1967)) shows the spectrogram for /d/ pronounced alongside various vowel sounds.
For the perception of /di/ there is a first formant (F1) transition, which involves a rapid rise from a low position until it reaches the position necessary for F1 of the /i/ vowel (this transition is a cue for all voiced stops). There is also a second formant (F2) transition, which moves from a position close to 1800 Hz to the location of the second formant in the /i/ vowel. The F2 transition differs in direction and the frequency region traversed when /d/ is pronounced alongside other vowels. The invariant cue thought to explain the homogeneous /d/ perception in all instances is the 1800 Hz locus of the origin of the second formant transition (Delattre and Cooper 1955, Liberman et al. 1967, 1957). Other voiced stop consonants (/g/ and /b/) exhibit similar second formant transitions and loci of origin cues (Ainsworth 1988).

Temporal information, such as the timing of onsets and length cues are also important to the perception of many speech sounds. Vowel perception involves a more limited range of cues, but the exact cues to even these simpler sounds, when heard in running speech and naturalistic settings, are not certain. For the reasons given below this chapter will focus on how mechanisms of timbral constancy and compensation affect the perception of vowel sounds:

1. It is necessary to limit the focus of this literature review to one type of sound. Vowels are very common in speech and also thought to be most sensitive to colouration because they are most reliant on spectral characteristics for their perception.

2. They are the simplest speech sounds. Fewer spectral cues and less time domain information is thought to be involved in their perception compared to other speech sounds. Isolated monophthongs can be perceived using only steady state spectral information. They can be synthetically reproduced with accuracy and
are more easily spliced from speech sections (Liberman et al. 1967). For this reason they are a good place to start an investigation into speech perception. Most of the research into invariant speech perception across talkers has examined these sounds.

3. They may be more similar to the majority of non-speech sounds because they may be perceived more continuously rather than categorically compared to other speech sounds (see Section 5.1.4).

5.1.2 Intrinsic processes—target and modified-target theories

Target theories of speech claim that the perception of all phonemes can be fully determined by a set of acoustical targets that need to be reached for perception to occur. For vowel perception ‘simple’ target theories (Strange 1989) have required that the appropriate F1 and F2 targets are produced (Blandon and Fant 1978, Carlson et al. 1975, Delattre et al. 1952, Fant 1960, 1978, Peterson 1952). The importance of these targets is seen in studies where recognisable vowels have been synthetically produced using resonators with filters set at F1 and F2 frequencies (Delattre et al. 1952, Fry et al. 1962, Helmholtz 1863). A number of studies using multidimensional scaling have demonstrated that the two most important dimensions of vowel perception are F1 and F2 (Carlson et al. 1975, Johnson 2005, Plomp 1975, Pols et al. 1969). ‘70% of the differences between vowels were explained by two orthogonal dimensions, representing differences in F1 and F2’ (Ainsworth 1988).

According to simple formant theories each vowel should occupy a unique position in 2 dimensional space where each axis is related to the frequencies of F1 and F2. However, Peterson and Barney (1952) showed variation in the position of F1 and F2 for American English isolated vowels produced by women, men and children (See Figure 5.2). When the same vowel is produced by different talkers, not only do formants vary around a single vowel target position but there is overlap in vowel space. This results in phonological identity, whereby the perception of the same vowel is cued by different formant frequencies for different talkers. Also, the same combinations of formants are perceived as different vowels when produced by different talkers. This vowel overlap
causes potential confusion. However, despite acoustical overlap, listeners perceive the intended vowel with high accuracy. This work neatly describes the impact of spectral colouration by the VT on the 2 most important dimensions of the speech signal for vowels but the problem may be larger than is shown here. Variation between talkers may also cause variation in other cues such as higher formants, which might also be relevant to perception, and overlap in respect of these. Perceptual invariance in spite of this physical variation illustrates a possible compensation mechanism that allows the listener to perceive each phoneme as if overlap does not occur.

**Gender and age differences**

Vocal tract differences between talkers are responsible for this variation in formant locations. VT differences are largely caused by gender and age. The anatomy of men's
and women's VTs differ. The VT of men is 30% larger than that of women and men have a longer pharynx and a larger pharynx length to mouth length ratio (Fant 2001). VT length can vary between genders from approximately 18 cm in adult males to 13 cm in adult females (Ainsworth 1988). Due to shorter vocal tracts women's formants are in general higher in frequency than those of men (Flynn 2011, Holmes and Holmes 2001). Formant centre frequencies for adults can vary by as much as 25% between men and women (Holmes and Holmes 2001, Lee and Rose 1998). The spectrogram in Figure 5.3 from Johnson (2005) shows variation in formants in the word 'cat' produced by a female and male talker. F0 also travels within a different range. F0 is about an 8ve higher for women compared to men and higher still for children (Holmes and Holmes 2001). The gender of young children can be recognised in speech. This is mainly cued by differences in frequencies of the vowel formants, particularly F2, rather than F0 (Johnson 2005).

![Figure 5.3: Taken from Johnson (2005). Spectrogram of a man and a woman saying 'cat'. The three lowest vowel formants are marked as F1, F2 and F3.](image)

**Formant ratio theories**

The work of Peterson and Barney (1952) shows the problem that VT variation presents for simple target theories of speech. Some researchers have used formant ratio theories to explain vowel perception claiming that it is the relative position of formants that is an important cue, not absolute values. The first report of a formant ratio theory of vowel perception is seen in the work of Lloyd (1890a,b). Lloyd declared that ‘like articulations produce like perceptions of vowel qualities, and like articulations produce like ratios of the formants’. Potter and Steinberg (1950) elaborated on this concept: in
vowel perception ‘a certain spatial pattern of stimulation on the basilar membrane may be identified as a given sound regardless of position along the membrane.’ This theory is inspired by analogy to musical chords ‘Musical chords, for example, are identified in this manner [...] the ear can identify a chord as a major triad, irrespective of its pitch position.’

The formant ratio theory is appealing because it is in accordance with Gestalt theory, an influential theory which describes the process of perception in all domains (Bregman 1990, Koffka 1935, Kohler 1929, Traunmüller 1981, 1984). Gestalt theory proposes that perception is based on the whole object and that perception of the whole comes before perception of the parts. This allows for the principle of invariance, whereby an object is perceived as the same object regardless of various transformations, including: rotation, scale and elastic deformations. It is well known that visual objects are perceived to be the same object regardless of the position of the image on the retina. A similar constancy despite position on the basilar membrane may apply to speech (Bregman 1990, Johnson 2005). Sussman (Sussman 1989, Sussman et al. 1997) proposes that combination sensitive neurons exist to determine formant patterns. This is similar to the feature detection mechanisms proposed to account for line and edge detection in vision (Hubel and Wiesel 1959, Shapley and Tolhurst 1973).

Formant ratio theories explain some of the variation in vowel formants that is to do with vocal tract variation and that might be caused by age and gender of the speaker (Chiba and Kajiyama 1941, Joos 1948, Peterson 1961, Potter and Steinberg 1950). This is illustrated in Figure 5.4. The same data as shown in Peterson and Barney (1952) is plotted but with the addition of lines to represent equal F1 and F2 ratios. It can be seen that relative position explains some of the variation between talkers but there is still variation around the specified ratios. Peterson (1961) showed that F2/F3 ratios reduced the ‘perceptual scatter’ further. However, some vowels remained difficult to classify.
Target theorists have maintained that the data of Peterson and Barney (1952) can be explained by the fact that not all of the relevant acoustical targets are specified in their research. Vowels are not fully transcribed by F1 and F2; other targets are important and may act by perceptually modulating the perceived frequencies of F1 and F2 or may simply add another dimension of variation that is relevant to identification.

If variation is in part due to age and gender then cues to this may be used in perception to normalise vowel targets. Miller (1953) examined the effect of fundamental frequency (F0) of speech on vowel categorisation. Miller (1953) proposed that knowledge of F0 may allow for automatic adjustment of F1 and F2 based on estimated vocal tract size. F0 is thought to be an *intrinsic* source (residing with the vowel) of information about talker variation (vocal tract length) (Nearey 1989). In Miller (1953) when the
F0 of synthetic vowels was doubled there was a shift in vowel categorisation. This is found to occur most for /u/, /æ/ and /ʌ/, the vowels centrally positioned in vowel space, which are also subject to the most acoustic overlap. Nusbaum and Morin (1992) and Halberstam and Raphael (2004) found that for mixed and single talker phoneme lists the presence of F0 in phonated speech, compared to whispered speech where F0 is absent, helped to disambiguate vowels where there was overlap. It has also been shown that large discrepancies between F0 and formants increase error ratings in vowel identification (Lehiste and Meltzer 1973, Peterson and Barney 1952).

However, it can be seen in the speech of the cartoon character Popeye (who has a high F0 and low formant range) or Julia Childs (who has a low F0 for a woman but high formants) that the correlation between tract length, F0, and higher formants is not reliable (Nearey 1989). The importance of F0 is also put into doubt by studies that show that perception is not greatly disrupted where vowels are whispered (F0 absent) (Rosner and Pickering 1994), and in sine wave speech without a fundamental (Remez et al. 1981), providing evidence that F0 is not critical to perception.

These findings are somewhat contrary to the source filter model of speech (Fant 1960), which claims that filter resonances are responsible for producing speech sounds and that the fundamental frequency or pitch of the voice does not have an effect on the objective or perceived location of formants. But it is acknowledged that there is not complete independence. Pitch may signal the listener to expect that the spacing of formants may be larger or shifted to a lower/higher range and the relationship between F0 and the rest of the signal must be more complex due to the convolution of a source that may change in spectral shape with pitch and a filter that changes in shape in a complex manner (Rothernberg 2008). The complexity of the relationship between F0 and the rest of the signal is likely to mean that any adjustment factor based on F0 is not simple. This may explain some of the mixed findings in this area. In spite of this, however, Miller (1989) claims that his data, along with a number of other studies (Fant et al. 1974, Fujisaki and Kawashima 1968) show that F0 can explain a significant proportion of the overlap seen in Peterson and Barney’s (1952) study.

‘It is well known that, under most conditions, the identity of a perceived vowel depends strongly on the formant values of the spectrum and is
independent of voice pitch. However, under certain circumstances, the voice pitch can influence a vowel’s identity, even when the formants are fixed’ (Miller 1989).

The higher formants may also be cues to vowel perception. Miller (1953) noted that F3 reduces variation in perception between talkers. F3 is thought to be a better estimator of vocal tract length as it varies little between vowels but by a large amount between talkers. F3 was found to explain the overlap between 3 vowels within the [u] region in Peterson and Barney’s (1952) study (Peterson 1961, Potter and Steinberg 1950). Fujisaki and Kawashima (1968) investigated the effect of F3 alongside F0 changes more fully. An F3 shift of 1500 Hz produced a vowel category boundary shift of 200 Hz in the F1-F2 space for a /u/-/e/ continuum. However, this was not replicated by Ainsworth (1976) in a similar experiment with English vowels. Some studies have averaged F2 and higher formants into single spectral prominence (F2, F3 (Syrdal and Gopal 1986), F2, F3, F4 (Blandon and Fant 1978, Carlson et al. 1975)) and have shown a reliable reduction in overlap for boundary frontal rounded vowels with this approach (Blandon and Fant 1978, Carlson et al. 1975). The consideration of F0, F3 and higher formants as reference cues to provide normalisation is an extension to the formant ratio theories. However, as mentioned previously, it remains the case that only a small amount of variation can be explained by using a limited number of reference cues. Further, the role of reference cues other than F0 and F3, if they contribute to normalisation, may be complex.

Transitions

Transitions in formants are important cues to perception. Transitions in formants do not occur for monophthongal vowel sounds but there is a transition into the vowel from flanking consonants (Strange et al. 1983), which can provide a cue to vowel identity (see Section 5.1.3). Further, even for isolated vowels there are small amounts of diphthongisation within the ‘steady state’ portion and this has been shown to improve identification (Hillenbrand et al. 1995). A linear time-invariant colouration by the transmission channel may affect the perception of change in formants less than the steady state spectral cues, so transitions may be useful to speech perception in light of
Strange et al. (1976) showed that Consonant Vowel Consonant (cVc) vowels were more accurately identified than isolated vowels. 31% of vowels were misidentified in isolation compared to 10% in consonantal context (see also Strange et al. (1979)). Assmann et al. (1982) found gated vowels (vowels with no consonants or transitions from consonants) were identified less accurately than full vowels with transitions remaining. Hillenbrand and Nearey (1999) also found that with synthesised steady state-formant vowels correct identification was 74% but improved to 89% with synthesized vowels with formant frequency trajectories included. Strange et al. (1983) and Nearey and Assmann (1986) showed that identification with silent centre stimuli in which the vowel centres and vowel duration cues were removed, leaving only transitions, was not significantly worse than isolated vowels and cVc stimuli. This suggests that when the steady state portion of the vowel is removed from the vowel nucleus identification is still possible. The studies appear to imply that transitions and spectral change are important cues in vowel recognition. If the information in transitions is important these studies might explain why cVc identification is often as good as isolated vowel identification despite problems of co-articulation, which is another source of distorting variation in the speech signal (see also Section 5.1.3).

However, Macchi (1980) found no improvement with consonantal context and provided evidence that the results in Strange et al. (1976) were due to experimental design factors. It appears that static cues (spectral and duration) and dynamic cues may both be sufficient to vowel identification but surprisingly neither of them appear to be necessary cues. These findings have led Miller (1989) and Hillenbrand and Nearey (1999) to state that vowel theories should not be thought of as prescribing points in normalized vowel space; a description of trajectories in vowel space is useful.

**Duration cues**

A reliance on temporal information in the face of spectral variation could aid perception. Generally temporal information is thought to be less important to the perception of vowels than other speech sounds but the duration of a vowel may affect
its perceived timbre. In American English the duration of the vowel is thought to be
the sole determinant of whether an ‘e’ as in bed or an ‘a’ as in trap is heard (Bennett
1968). Bennett (1968) found that duration is an important cue when vowels are close
together in spectral space for English and German vowels. Ainsworth (1972) also found
that duration effects are most prominent if the vowel is in the centre of F1 and F2 space,
where it is more easily confused.

According to Strange et al. (1983) both reducing and lengthening duration reduced
identification accuracy when vowels were present without consonants. Also, when
duration was fixed to be the same for all vowels identification accuracy was reduced.
The authors explain that making the steady state vowel a single length may have
also removed all within-vowel movement (i.e. any diphthongisation). Rather than the
duration itself being a cue to the vowel sound the transition or the rate of transition
into and out of the vowel again appears to be important. Strange et al. (1983) found
that with cVc syllables, there was no effect of reducing the duration of a silent vowel (a
vowel with the centre portion removed leaving only the consonants and transitions into
and out of the vowel). Comparing this result with other studies, where duration of the
vowel (where actually present) is found to be a relevant cue in cVc vowels. This may
indicate that the length of time that the steady state portion of the vowel is actually
present is a cue rather than merely the elapsed time between consonants but this can
be traded off against transitions. This may be because there is no novel information
in the silent vowel but it is possible that the spectral change during the vowel steady
state contributes to identification.

Intrinsic processes summary and discussion

Target models of speech perception propose that when various acoustical parameters
are produced the perception of a vowel will be formed. Targets may be formant
positions, formant ratios, dynamic spectral patterns or duration cues. One means
by which speech may maintain stability across channels is by the use of targets that
are spectral magnitude peaks (formants) which may be more robust to variation caused
by transmission channels compared to wideband spectral cues due to their high signal
to noise ratio. However, formants have been shown not to be immune to variation
caused by vocal tract differences. Due to different limitations of the VT between talkers, canonical formant targets cannot be reached so the location of formants is shifted when different talkers aim to produce the same vowel. Therefore, it cannot be said that formants produce stability when the VT is the channel. The extent to which other channels such as loudspeakers and rooms can influence the location of formants or other peaks in the source is not clear. It is likely that their effects in relation to shifting formants will be less extreme than that of the VT but disruption in the location of formants may still occur due to the limitations of those systems. Distinctive resonance peaks or formants are distinguishing features in speech and play a particularly important role in speech identification. Formants are not known to occur to the same degree in non-speech sounds and may not play such an important role in timbre and the recognition of non-speech auditory objects. For these sounds the wideband spectral shape appears to be more relevant to timbre. Because of a lack of reliance on resonance peaks, which have a higher signal to noise ratio, non-speech timbre may be less stable across channels such as loudspeakers and rooms. One means by which speech may maintain stability across channels is by the use of targets that are spectral magnitude peaks (formants) which may be more robust to variation caused by transmission channels compared to wideband spectral cues due to their high signal to noise ratio. However, formants have been shown not to be immune to variation caused by vocal tract differences. Due to different limitations of the VT between talkers, canonical formant targets cannot be reached so the location of formants is shifted when different talkers aim to produce the same vowel. Therefore, it cannot be said that formants produce stability when the VT is the channel. The extent to which other channels such as loudspeakers and rooms can influence the location of formants or other peaks in the source is not clear. It is likely that their effects in relation to shifting formants will be less extreme than that of the VT but disruption in the location of formants may still occur due to the limitations of those systems. Formants are distinguishing features in speech and play a particularly important role in speech identification. Formants are not known to occur to the same degree in non-speech sounds and may not play such an important role in timbre and the recognition of non-speech auditory objects. For these sounds the wideband spectral shape appears to be more relevant to timbre. Because of a lack of reliance on resonance peaks, which have a higher signal to noise ratio, non-speech timbre may be less stable across channels
such as loudspeakers and rooms. Further, it should be noted that any mechanisms that reduce the spectral characteristics of the VT do not do this to the extent that all variation in spectrum between talkers is reduced. Some variation between talkers that does not cause overlap may be necessary to identify the talker and compensation for VT characteristics is not expected to be complete. If these mechanisms compensate for other channels it is likely that compensation for these channels will also not be complete.

Listeners appear insensitive to variations in formants caused by the VT. Mechanisms behind this may be redundancy in the speech signal. Where there are many cues that are sufficient to signal a speech sound, and only some of these are distorted by the channel, there exists a redundancy of information that can account for perceptual constancy. Purely temporal factors such as duration and onset time are not affected by spectral colouration. However, alone, temporal cues play a relatively small role in vowel perception and in the perception of other phonemes. Dynamic cues (time varying changes in spectral cues) appear to be particularly important to speech. If spectral change is more perceptible against a time-invariant colouration then this may explain the accurate perception of many speech sounds. However, colouration by the channel may affect the range over which transitions travel or their locus of origin/endpoint, so these cues are also affected by colouration. Further, temporal envelope distortion by the channel is also likely to occur alongside spectral envelope distortion. Spectral components are key aspects of speech. Therefore, it cannot be said that the listener can often turn to non-spectral cues where there is channel colouration. The static spectrum still appears to be a primary cue to recognition for some sounds (most notably vowels) and accurate perception of these sounds in light of channel colouration remains to be explained. It is possible that the transitions into vowels from the flanking constants may have a role in ensuring that colouration is not a contributor to perception by weighting perception towards perceiving change rather than steady-state values.

It is clear that a target theory of speech is too simplistic. Targets are not uniform across talkers. ‘modified target models’ can explain some of the variation in formants in identically perceived vowels: targets modified in light of one or two others (i.e. normalised targets) explain a little variation, but the use of these targets is unreliable. Further, research has only attempted to explain variation in just 1 or 2 degrees of
freedom such as VT length, whereas variation between talkers’ VTs is complex. The F0 and F3 and F1/F2 ratios do not describe the whole impulse response of the system.

5.1.3 Extrinsic processes—adapting using spectral and temporal context

The methods described above are *vowel intrinsic* and only consider information within the vowel when attempting to normalise speech cues across speakers (Nearey 1989). Within-vowel information may not be enough to describe the vocal tract area function. This section describes how extrinsic information from outside the vowel may be used in normalisation, specifically information gained from the talker in their speech leading up to the vowel. Further, an added source of variation to the spectrum of running speech is described. This is variation is caused by the phonetic context. The picture of stable perception is therefore more complex than the examination of isolated speech segments suggests.

Joos (1948) is thought to have first put forward the notion that a listener uses information about a speaker’s formant frequency range obtained from hearing them pronounce other words: ‘unknown vowels are identified in terms of the way in which their acoustic structure fits into the pattern of sound that the listener has been able to observe’ (Joos 1948). This information about the speaker’s unique speech cues may further aid perception. Ladefoged and Broadbent (1957) investigated this hypothesis in a study which showed that the identification of vowels is influenced by prior experience with the speaker’s ‘formant range’ by listening to a sentence of their prior speech. A synthesized precursor phrase, ‘please say what this word is’ was followed by a *bVt* test word. The vowel in the middle of the test word was varied in F1 so that it sounded between ‘bit’ (low F1) and ‘bet’ (high F1). The F1s of all phonemes in the precursor sentence were also either all lowered or all raised so that the phrase sounded like it was spoken by a different talker. When the F1 in the precursor was lowered the test word was reported to sound like ‘bet’ (high F1). When the F1 in the precursor was raised the test word sounded like ‘bit’ (low F1). This result demonstrates a shift in the vowel category boundary (the point at which vowel identification changes) in the opposite
direction of the manipulation to the precursor. Similar effects were seen for alteration of the 2nd formant. The effect was reduced significantly with a 10 second gap between test sound and precursor, indicating that the effect of the precursor sentence was short-lived. The authors observed that where the speaker’s F1 and F2 was shifted further apart, the test vowel was reported to be perceived as a vowel having more narrowly spaced formants and that listeners calibrate vowel categorisation to the formant range of the speaker. They explain that context of the precursor sentence provides a coordinate system within which to judge and identify the test vowels. They interpreted this result within the framework of adaptation level theory which assumes that perceivers regularly gauge the range of a stimulus continuum in the process of formulation of psychophysical judgements (Helson 1948). They claim that this is a form of relational processing whereby stimuli are perceived in context and it is their relationship to the context that is relevant rather than their absolute value. This process appears to maintain perceptions of speech targets in a relatively central perceptual position across different spectral contexts. In terms of Peterson and Barney’s vowel space (Peterson 1952), a speaker with a high range would produce a phoneme with higher F1/F2 locations but the listener would hear this phoneme within running speech as being at the midpoint rather than high. This is indicated by the fact that when presented with actual midpoint format values these are heard as being relatively low (they are heard as low vowel, not the midpoint). Ladefoged and Broadbent (1957) noted that the dependence of vowel perception on context is similar to that seen with colour perception:

‘It is obvious that this experiment provides a demonstration of perceptual constancy in the auditory field; that is an auditory phenomenon somewhat parallel to the visual case in which the response evoked by a stimulus is influenced by the stimuli with which it is closely associated. An example is the correct identification of the colour of an object in widely differing illuminations. Consequently it is hoped that further investigation of the auditory phenomenon will provide data which are of general psychological interest’.

Dechovitz (1977) confirmed the findings of Ladefoged and Broadbent (1957) in a similar study using natural speech. The formant range in a precursor sentence, ‘please say
(bVt) for me’ was altered by using adult and child male voices produced at the same F0 and rate (the difference between precursors being only in formant frequencies). The listener’s task was to identify the bVt test syllable produced by an adult male. When the child precursor sentence was used there was a 54% increase in misidentification. Confusions were found to be large for ‘bet’ and ‘bat’ test syllables. The formants in these words are close to that of ‘bat’ and ‘but’ of a child and there was a tendency to perceive them as such. The authors state that ‘these results suggest that listeners adjusted their criteria for vowel identification according to formant ranges specified by context material’.

Other evidence that experience with a speaker’s vowel space is useful for vowel identification is mixed. The recognition of vowels in mixed talker versus single talker lists has been examined. These studies expect that experience with the same talker should increase identification thorough exposure to the speaker’s vowel space. In Strange et al. (1976) increased accuracy of identification occurred for single talker lists of isolated vowels (31.2% error) compared to mixed talker lists (42.6% error) and single talker lists of pVp vowels (9.5% error) compared to mixed talker lists (17.0% error). Assmann et al. (1982) showed a small but significant improvement for blocked (4.09% error) versus mixed (5.43% error) speaker lists of isolated vowels, and gated isolated vowels, where transitional information and duration is removed (blocked 9.50% error; mixed 13.75% error) but noted that error was so low for mixed vowels that speaker experience is likely not to be important for identification. Mulennix et al. (1989) found advantages for single speaker conditions over a range of natural words. In Verbrugge (1976), in a test using pVp test stimuli, identification for single talker lists was improved. The most confusable vowels benefited most from single talker presentation. Again error was low with mixed talkers (17%) leading the authors to assert that ‘there is clearly a great deal of information within a single syllable which specifies the identity of its vowel nucleus’. Verbrugge (1976) declared that ‘speaker variation is not a perceptual problem, contrary to Ladefoged and Broadbent (1957)’. The mixed findings point to a small effect of speaker experience that suggests that ‘extreme theories of speaker normalisation of [spectral] range must be rejected’ (Assmann et al. 1982). However, an issue with these studies is that it is not always clear the extent to which the stimuli measured come from regions of overlap, where
experience listening to the same talker is likely to be most beneficial.

The question remains as to what cues are being used in the adaptation to the talker seen in the above studies. Liberman (1973) proposes that if spectral range is relevant the *point vowels* (/I/, /a/, /u/) should be the primary calibrators of vowel space. Point vowels represent the extremes of articulator space and formant frequency values and are reported to be the only vowels which are not ambiguous until calibration (because they are related to a unique vocal tract area function (Lindblom and Sundberg 1969)). According to Ladefoged and Broadbent (1957) experience with a speaker’s point vowels should enhance the identification of test vowels. However, when Verbrugge (1976) presented 3 point vowel $kVp$ precursors followed by a $hVd$ syllable test stimulus, the error rate was 12.2% for no precursor stimuli and 12.9% with the precursor stimuli. Even the most ambiguous vowels were not reliably aided by point vowel precursors. In favour of the utility of point vowel precursors Gerstman (1968) has successfully classified the data of Peterson and Barney (1952) using a method that involved calibration based on point vowels. His method prescribes the use of point vowels to work out a spectral centre as a vocal tract reference. It is therefore possible that in this case the spectral information gleaned from the speaker, is to do with a spectral average rather than specific formant range values. This might represent a different process of normalisation than proposed by theories that state that a sample of the spectral range is useful. Assmann et al. (1982) claimed that an increase in performance in single talker conditions might be to do with the dynamic information that is conveyed by experience with the talker rather than experience with their formant range. However, in this study there was increased recognition for gated vowels (where transitional information had been removed) with the single speaker condition (9.5% error versus 13.6%) which indicated that a single speaker condition provides some useful experiences with only ‘steady state’ components of the vowels.

Studies that have reported benefits of identification for single talker lists have described adjusting for speaker differences as an *active process* and cognitive processes, rather than automatic peripheral processes are more commonly suggested as mechanisms by which adjustment to the talker occurs. Increased activity within central speech processing areas of the cortex has been shown for mixed talker presentations (Wong et al. 2004). In Nusbaum and Morin (1992) participants were asked to remember a
series of numbers. Reaction time was increased and recall performance was poorer in the mixed voice condition. The authors commented that normalisation is controlled by a contextual-tuning process: ‘attentional demands are increased (in the mixed voice condition) because of the presence of variability. Speech perception requires active processing to reduce the set of possible responses to a single response’. Summerfield and Haggard (1975) also showed longer latencies for categorizing synthetic vowels when target items were preceded by syllables with different synthetic voices. The authors suggest that ‘the increase in response time to talker variability reflected additional processing needed for vocal tract normalisation’. However, lower level processing was suggested by Mulennix et al. (1989). When using a wider vocabulary of words, rather than nonsense syllables, Mulennix et al. (1989) found increased recognition and faster reaction times for single talker lists compared to mixed talker lists. Lexical density (the number of words that differ by only 1 phoneme) and word frequency (how common the word is in the English language) also result in differing identification accuracy (high-density words and less frequent words are more difficult to identify). Mulennix et al. (1989) did not find a reliable interaction between speaker condition (mixed versus single) and lexical density or frequency and claimed that this indicated that mixed talker effects work at lower levels of word processing than these post-lexical factors. There was, however, an interaction between signal degradation and speaker condition with degradation effects resulting in even poorer identification with mixed talker lists. This finding indicates that talker compensation effects occur at the same level of processing as the analysis of signal degradation. Mulennix et al. (1989) reports that:

‘When the processing of low level cues in the signal becomes disrupted as a result of signal degradation the effect of talker variability on perception becomes greater. This finding is consistent with a view suggesting that talker normalisation processes invoked by changes in the talker voice from trial to trial are intimately related to early processes involved in encoding the sensory input into a phonetic representation.’
Phonetic context and co-articulation

With carefully produced slow speech it is often possible to extract a steady state portion with identifiable formant frequencies from the nucleus of the vowel (Assmann et al. 1982, Macchi 1980). However, in normal rapid speech such portions frequently do not occur (Lindblom and Studdert-Kennedy 1967, Verbrugge et al. 1976). The syllable is not discretely partitioned and the information contained in the vowel nucleus varies depending on the flanking consonants (Liberman et al. 1967). This occurs because the speed at which articulators move is limited and there is not usually sufficient time for a steady vowel configuration to be reached before the shape of the vocal tract must be changed. This process is referred to as co-articulation. There is an assimilation of adjacent sounds, whereby target undershoot occurs: the acoustical targets for each adjacent sound are not reached but a position between the two targets is reached instead (Lindblom 1963). The effect of undershoot is larger with faster speech rates (Lindblom 1963, Lindblom and Studdert-Kennedy 1967). As well as assimilation there is a blurring of acoustic cues across adjacent speech sounds so that some of the acoustical information of a phoneme is smeared across adjacent phonemes, each phoneme contains information about itself and also about its flanking sounds. Undershoot causes overlap in vowel space similar to that caused by speaker variation shown by Peterson and Barney (1952) (Lindblom and Studdert-Kennedy 1967). However, this assimilation and smearing due to co-articulation is not perceived. Phonemes and syllables are heard as if they can be easily partitioned and without overlap. Target theories alone do not account for this (House and Fairbanks 1953). Compensation for the effects of the phonetic context appears to be necessary to combat the effects of undershoot caused by co-articulation. The processes behind this appear to be similar to those involved in adjusting to the VT characteristics of the talker and may indicate mechanisms for channel compensation more generally.

Consonantal context effects

The consonants that surround vowels have the effect of degrading the formant structure of the vowel nucleus via undershoot (Nearey 1989). However, it was also seen in Section 5.1.2 that consonantal context may have a beneficial, as well as detrimental, effect on
perception as there is evidence that vowels within cVc contexts can be recognized at least as well as vowels in isolation (Macchi 1980, Strange et al. 1979, Verbrugge et al. 1976). An important study by Lindblom and Studdert-Kennedy (1967), demonstrated the effect of consonantal context on vowels by simultaneously examining the acoustic properties and perception of wVw and jVj syllables. The vowel token in the middle of these syllables varied along a /i/ (higher formants) to /u/ (lower formants) continuum in 20 steps. It was expected that these vowels would be perceived to be the same in both /w/ and /j/ contexts if consonantal context had no influence on the vowel. If consonantal context has an effect, vowel acoustics and identification was expected to vary with flanking consonants. The formant frequencies in /w/ start from a lower formant position than u/i and, due to undershoot, the formants in the vowel section were expected to be lowered when presented within a wVw context. The formants in /j/ start from a higher position than those in the vowel tokens and undershoot was expected to raise the formants in the /j/ flanked vowel compared to unflanked vowels. This objective effect was found. However, compensation for this effect was also shown with the categorization boundary between /u/ and /i/ shifting towards the higher end of frequency scale for /j/ consonants, so that more /u/ vowels are perceived compared to the number that would be expected with an undershoot effect or if the vowel was unflanked. The categorisation boundary shifted towards the lower end of the frequency scale with w consonants so that fewer /u/s were perceived. Thus compensation for undershoot was demonstrated. This compensation effect has been named perceptual ‘overshoot’ by the authors. The overshoot effect appears to have the effect of increasing spectral differences between adjacent sounds and may result in an expanded vowel space. Johnson (2000) found a similar effect when he asked listeners to label prototypical vowels. He observed that listeners identify vowels as if there is a perceptually expanded vowel space when listening to vowels in consonantal context. He calls this process ‘adaptive dispersion’.

A number of studies following Lindblom and Studdert-Kennedy (1967) have shown the perceptual overshoot effect with different cVc syllables (Holt 1999, Holt et al. 2000, Mann 1980, Nearey 1989) (some of these studies are discussed in relation to non-speech perception in Section 5.3). For example, Mann (1980) found that listeners were more likely to perceive test sounds from a ga-da continuum, which varies in F3 between two
phonemes, as /ga/ (low-frequency F3 energy) when preceded by /al/ (high frequency F3 energy) and as /da/ (high frequency energy) when preceded by /ar/ (low-frequency F3). It was concluded that this is evidence of a perceptual overshoot effect.

It has been suggested that the overshoot effect is a result of automatic speech-specific processes that require knowledge of articulatory movements, such as the *gestural account* of speech perception, which is similar to the motor theory of speech perception, whereby listeners can deduce the intended phoneme through the perception of articulatory movements attempting to produce that phoneme (Diehl et al. 2004, Fowler 1990, 2006, Mann 1986). However, general auditory processes such as cognitive relational processing (as suggested by Ladefoged and Broadbent (1957) to explain VT compensation) have been suggested by Lindblom and Studdert-Kennedy (1967) themselves as an explanation for overshoot. Further, Lindblom and Studdert-Kennedy (1967) also propose that neural adaptation in the auditory periphery might explain overshoot. This would affect all sounds entering the ear. A general auditory cause has since been supported by Lotto et al. (1997), Lotto and Kluender (1998), Holt (1999), Holt et al. (2000) and Holt et al. (2005) in studies using non-speech sounds. These studies are discussed further in Section 5.3. There is growing evidence to suggest that the effect may not be specific to speech.

**Duration and rhythmic context**

The importance of vowel duration and rate of transition in to and out of the vowels on vowel identification has been discussed above (see Section 5.1.2). Context effects seen in speech perception may be due to the effect of the temporal context on temporal cues. Ainsworth (1972, 1974) provided evidence of the influence of rhythmic context on vowel identity. A cVc test word (hid, had, head, heard and hud) was played with a vowel duration of either 120, 240 or 480 ms. A precursor sequence of vowel followed by silence (times 3) was also played. The lengths of vowels and silence were equal, and varied between 120 and 600 ms. The number of long and short vowels heard in the test stimulus was found to depend on the rate of precursor sequence. When the precursor rhythm increased, the vowel in the test stimulus was perceived to be longer; when the precursor rhythm decreased, the vowel in the test stimulus was perceived to be shorter.
This effect was only reliable when the duration of the test vowel was in the ambiguous region (240 ms) and when the test vowel consisted of ambiguous formant frequencies. The effect therefore may be said to contribute to the identification of vowels where formant overlap occurs.

Verbrugge et al. (1976) also noted that prior experience with a speaker may enhance vowel perception by providing information about rate and stress of speech and the duration of the vowels, rather than by providing information about the speaker’s formant range. Listeners heard a \( pVp \) stimulus with all of the nine monophthongs. The syllables were either spoken slowly in citation form, or they had been extracted from a sentence spoken at a normal rate. It was hypothesised that vowel identification would be more difficult for vowels extracted from running speech because the syllable would be shorter, with no steady state. Notable undershoot was seen in the extracted speech and there was a 24% error rate compared with 17% error for citation vowels. When listening to extracted syllables listeners were biased to hearing shorter vowels than they would hear when the vowel remained within the running speech. This indicated that they heard the extracted syllables as if they had been spoken in isolation and more slowly, like the citation syllables. The authors claimed that this was a strong indication that information about talkers’ tempo is critical to achieving constancy. Two further conditions involved preceding the test stimuli with with a train of point-vowel precursor’s hi/ha/hu spoken at a slow rate, and alternatively the \( pVp \) presented in original context: ‘the little \( pVp \) chair is red’. Identification of the extracted vowel showed further increased error with point-vowel precursors compared to hearing the vowel in its original context sentence. It was concluded that point vowel precursors enhanced the perception of a pattern of slow citation type speech, increasing the perception that the extracted syllables were also produced in this fashion. When test syllables where produced within the original context sentence, no significant identification errors occurred. Error rate was the same as for the original carefully spoken citation syllables. The authors declared that the original sentence precursor contained sufficient tempo information to allow for proper adjustment and that the results show that it does not matter if a steady portion of the vowel is not reached in fast speech as long as tempo information is available: ‘if the listener is tuned to ongoing tempo a short de-stressed syllable is as fully determined as a citation form
syllable’. The point vowel precursors would have provided spectral information but they did not reduce ambiguity in this experiment. The authors suggested that carrier sentences are more useful for defining tempo and stress, than spectral range for a given speaker. Dechovitz (1977) and (Miller and Eimas 1994) found similar effects.

Extrinsic processes summary and discussion

The research in this subsection shows context affects perception. Prior speech by the same talker perceptually shifts vowel identification in a contrastive manner (Ladefoged and Broadbent 1957). This effect suggests perception in light of the prior context and compensation for the frequency range of the context speech and for VT differences between talkers. The authors explain the effect as relational processing: listeners hear the target in light of the context of prior speech, with the context representing the perceptual midpoint, the perception of vowels or other phonemes is normalised using the context speech. This process appears to maintain perceptions of speech targets in a relatively central perceptual position across different spectral contexts. Relational processing was put forward as an explanation for the effect seen in Ladefoged and Broadbent (1957). However, further explanation of the mechanisms behind this were not given. Relational processing implies a cognitive mechanism and descriptions of an active process also imply deliberative thought and cognitive mechanisms rather than a quick automatic process like intrinsic normalisation. The time course is not well established. Ladefoged and Broadbent (1957) show an effect which comes about after listening to a sentence and which recovers with 10 s of silence.

It should be noted that this compensation for the spectral effects of the VT is far from certain. The effect seen in Ladefoged and Broadbent (1957) does not translate to consistent improvement in single talker lists (it is possible that information in a cVc test word provides all the necessary context). The point vowels, should reliably assist in the identification of a target vowel if prior information about the spectral range of the talker is utilised but there is no consistent evidence of this. Where improvement does occur there has been evidence to suggest that it may be due to temporal context rather than spectral. The most reliable effect of speaker experience appears to be to do with speech rate rather than spectral information. However, some studies that have
examined the effect of experience with the speaker’s spectrum have controlled for speech rate, so spectral information might offer a unique contribution to the disambiguating vowels. Further, because vowel identification is reasonably good with isolated vowels, it can be said that experience is not necessary for accurate identification of vowels. There must be adequate intrinsic information for this to be possible, or the vowels measured are not those with formants in the region of overlap.

The research in this section shows that immediately adjacent context speech (e.g. flanking consonants) also has the potential to disrupt speech perception because of undershoot. Undershoot arises due to limitations of the physical system similar to those relating to variation in VT shape across talkers. In both cases spectral targets are moved from canonical targets in different ways due to different limitations of the mechanical system. In the case of undershoot this is due to sluggish movement of articulators. This results in acoustical targets not being reached (Lindblom and Studdert-Kennedy 1967) and overlap in vowel space occurring. However, undershoot and the resulting overlap is not a problem for speech perception because perceptual overshoot occurs to compensate for assimilation. This results in the perception of the formants of the target vowel being shifted away from those of the flanking consonant. Overshoot may also be said to enhance differences between adjacent stimuli (Lindblom and Studdert-Kennedy 1967). This process is similar to that seen in Ladefoged and Broadbent’s (1957) study, and may be the result of the same or a similar mechanism. In both cases there is a contrastive shift in perception between a prior speech segment and a test sound. Both seem to involve adaptation to the spectral range or spectral average of prior speech. Like compensation for the VT characteristics, overshoot appears to involve hearing the target vowel in relation to prior speech and relational processing has been given as an explanation. Relational processing does not fully describe the mechanism but Lindblom and Studdert-Kennedy (1967) also give a more specific suggestion: peripheral neural adaptation. Low-level processes may be the cause of relational processing seen in the overshoot effect and in Ladefoged and Broadbent’s (1957) study of VT compensation. There is evidence that benefits from listening to a single talker occur at the same level at which signal degradation is processed rather than at a higher lexical level. Potential mechanisms of overshoot and the compensation for the VT observed by Ladefoged and Broadbent (1957) are discussed further in sections
5.2 and 5.3. It may be that the same underlying mechanism causes both Ladefoged and Broadbent's (1957) effect and overshoot. However, there appears to be a time course difference: the overshoot effect occurs with listening to a prior phoneme rather than sentence length precursors used in Ladefoged and Broadbent’s (1957) study. It appears to show compensation with a much reduced experience of the context compared to that study, which suggests a different mechanism may be involved in the two effects. Recovery of overshoot effect may also be shorter than 10s but this was not tested.

Overall, speech heard in context appears to be perceived accurately despite problems of co-articulation. Context allows for information within flanking speech sounds to contribute to identification and flanking sounds also produce overshoot which may allow speech sounds to be heard in the appropriate context (in relation to previous sounds made by the speaker) and this may further enhance contrast between adjacent sounds, compared to objective values. Differences between overshoot and compensation for VT variation may only be in the time frame over which the context is sampled: the preceding phoneme or the preceding sentence, and the aims of the process—removing the channel (the VT) and removing undershoot/enhancing spectral change between phonemes.

### 5.1.4 Categorical perception and categorisation

A set of targets that fully prescribes vowel perception across speakers has been elusive even after normalisation for speaker characteristics and phonetic context. Other approaches to phoneme recognition have been examined. Categorisation involves exemplar-based coding and the use of all the data in the acoustical signal (i.e. a very large number or targets) to statistically match the signal to stored canonical representations or exemplars of speech sounds. This categorisation may mean that small variations such as those caused by transmission channel distortions do not affect listeners’ ability to identify and categorise auditory objects. Categorical Perception (CP) is another process that might reduce variation caused by transmission channels. CP and its moderated version, the perceptual magnet effect, involves a perceptual reduction of acoustical data so that the larger variations between different auditory objects are easily perceived but small variations within object categories are not. CP
may be regarded as an intermediate position between target theories and categorisation because it involves segregation of phonemes based on changes in specific target features (e.g. voice onset time and format transition durations), but also the formation and storage of robust categories of speech sounds.

**Categorical perception in speech**

The theory of CP has developed in response to the observation that, acoustically, speech consists of continuous unsegmented sequences and yet each phoneme is perceived as a discrete, segmented, entity. It is also a theory of minimisation of variation in the speech signal and although not previously specifically examined as such, it may be a mechanism of compensation for VT variation, undershoot, and colouration caused by other channels. CP proposes that phonemes are represented in perception as being members of unique categories and once a categorical representation is formed finer distinctions between phonemes within the same category are no longer accessible. CP states that only variations in a sound large enough to result in the perception of a different category of object are perceived. For example a /b/ sound is represented as a /b/ and any variations in the signal will not affect this perception, until they are so large that they specify a different category of sound (e.g. a /d/).

CP is demonstrated most clearly with the perception of consonants. When a series of consonants that vary continuously in a distinguishing parameter (such as F2 transition) are presented, the listener tends to hear one consonant until a certain point in the series is reached when they will hear another. According to categorical perception the sounds are heard as belonging to one phoneme category or the other and never partly to both. The point at which perception changes is the 'perceptual boundary' or 'category boundary' (Ainsworth 1988). Liberman et al. (1957) demonstrated this effect for plosives /b/ /d/ and /g/. In an identification task listeners were presented with synthesised syllables which differed only in 2nd formant transition locus of origin. Below 1500 Hz /b/ was heard, between 1500 and 2000 Hz /d/ was heard and above 2000 Hz /g/ was heard (see figure Figure 5.5). There was a distinct change in category naming at certain points in the variation of F2. Also, in an ABX discrimination test, peaks in discrimination performance occurred when comparison stimuli were at
different sides of a category boundary. Listeners could not discriminate between stimuli where F2 varied within each category boundary. Both category boundaries and peaks in discrimination for stimuli across boundary are necessary for object perception to be considered categorical (Harnad 1987). For strong CP the ability to discriminate between two categories, where the stimuli do not cross a category boundary, should be impaired so that performance is at chance.

![Figure 5.5: Categorical perception for consonants. F2 locus of origin below 1500 Hz (stimuli 1-5) results in /b/ identifications, between 1500 and 2000 Hz (stimuli 6-10) results in /d/, and above 2000 Hz (stimuli 11-14) /g/ is heard. Taken from (Liberman et al. 1957).](image)

**Moderated CP—the perceptual magnet effect**

A number of studies in speech perception have shown that perception is not as categorical as was at first proposed. There is evidence that richness in the acoustical signal is still perceived and that category boundaries can be moved through learning and training. CP does not occur, or only occurs to a limited extent, for vowels (Feldman and Griffiths 2007, Fry et al. 1962, Liberman et al. 1957, Pisoni 1973, Repp 1981, Thyer et al. 2000) listeners can hear more variation than they can identify different vowel types. Cutting et al. (1976) found identification and discrimination curves similar to CP for vowels but this was less marked than for stop consonants. At no point was within-category discrimination at chance. Only consonants show strict CP. A moderated version of CP has been proposed by Kuhl et al. (1992) and the mechanisms behind this may be different to CP. Kuhl et al. showed that gradual warping of the signal occurs though a ‘Perceptual Magnet Effect’ (PME). Within categories finer
distinctions can be made for stimuli nearer to category boundaries, with decreased
discrimination at the nucleus of the phoneme where a prototype (a perfect example
or ‘good instance’ of the category) occurs (Rosch 1975). When comparisons of a
stimulus within a phoneme category are made against a prototype, all instances of
the stimuli appear to be more similar to the prototype. This effect does not occur
when a non-prototype stimulus is used for comparisons. A larger distance is required
between a prototype and its variants for them to be heard as the same distance between
a non-prototype and its variants. This effect demonstrates a concertina effect in
perception whereby variance is minimised within categories and maximised between
categories. The effect has been demonstrated in the English /i/ category, the German
/i/ category, and the Swedish /y/ category (Kuhl and Iverson 1995). The effect has
not occurred reliably for other English vowels. It has been found for consonants such
as the English /r/ and /l/ prototypes in English speakers amongst other continua.
However, evidence of no CP or PME for vowels comes from Stevens et al. (1969) who
found that there was no categorical perception for rounding in vowels in American or
Swedish listeners, although in Sweden there is a phonemic distinction between rounded
and un-rounded vowels.

Mechanisms of CP/the perceptual magnet effect

The effect of CP and PME is still a subject of debate and mechanisms for the PME
may be different from those behind CP. PME is thought to result from exposure to
and learning of a language which distorts the perpetual space underlying phonetic
perception. PME is shown for infants, occurring for native vowels at 6 months, but
not monkeys (Kuhl 1991). The learning process may enhance invariance for sounds of
the same linguistic category within a listener’s native language. The finding of PME
for English /r/ and /l/ prototypes in English speakers but not Japanese speakers, for
whom a distinction between /r/ and /l/ does not exist (Feldman and Griffiths 2007,
Miyawaki et al. 1975), suggests a learning effect. The categorical distinctions seen with
CP have been found to occur with younger infants (before language learning), however,
indicating that it is more likely to be innate and possibly a result of a lower-level
physiological process.
Because of the dissimilarity between vowel and consonant perception two different modes of perception have been proposed: ‘a categorical or phonetic mode and a continuous or auditory mode’ (Liberman et al. 1967, Stevens et al. 1969, Studdert-Kennedy 1973, Studdert-Kennedy et al. 1972). The difference between these modes is not well understood but may be to do with ‘the degree to which separate auditory and phonetic memory components are employed in the decision process during discrimination’ (Fujisaki and Kawashima 1969, 1970, Pisoni 1973). Fujisaki and Kawashima’s model proposes that when phonemes from different categories are presented the features of stimuli that are used to discriminate between the two phonemes are stored in ‘phonetic short-term memory’. Where the sounds are categorised as different, further discrimination of phonemes is not possible. Once two phonemes have been categorised as the same, however, the listener can use the stored auditory information about the fine acoustic detail preserved in ‘auditory short-term memory’ (also known as auditory sensory memory). The listener can then make a comparative judgement rather than an absolute one.

Evidence for this memory model of CP/PME was seen in Pisoni (1973) where a delayed comparison task was used to examine memory processes involved in judging categorically different and similar stimuli. ‘Same’/‘different’ judgements were made for within and between-category pairs of: bilabial stop consonants (/b/ and /p/), voiced stop consonants /b/ and /d/ and long and sort versions of steady state vowels /i/ and /I/. There were 0-2 second time gaps between comparisons. Results showed that between-boundary discriminations were generally good and not affected by time for either vowels or consonants, but some lessening of discrimination with time did occur for vowels. This result suggests that categorically different phonemes are stored in a memory store which stores a robust representation. This has been classed as a phonetic storage and this storage appears to occur whenever distinct phonemes are perceived. Within-category discrimination for consonants was low and not related to the time between comparisons. It appears that where consonants are from different categories they move into categorical storage but where they belong to the same category any perception of difference is lost, preventing fine discrimination between two consonants of the same category. Within-category discrimination was better for vowels. It was at maximum at .25 seconds and decreased with time between
5.1 LITERATURE REVIEW PART 1: COMPENSATION FOR THE VOCAL TRACT AND PHONETIC CONTEXT

comparisons. Within-category discrimination was even better for long vowels. The study shows that there may be a separate auditory short-term memory which facilitates vowel discrimination within categories but does not occur for consonant sounds. These smaller variations are only stored for a limited time however, and are not as robust as between-category representations. The authors reported that ‘it is clear from the results [...] that an increase in the delay interval affects vowel discrimination accuracy much more than consonant discrimination. Moreover the difference appears to be most pronounced for within-category vowel comparisons [...] the results strongly support the claim that differences in discrimination between consonants and vowels are related to differential usage of auditory short term memory’.

Fujisaki and Kawashima (1970) claim that the difference between vowels and consonants is because of cue duration. The cues differentiating consonants (VOT, transitions) are short cues, which cannot be remembered as well as vowel cues (steady formant frequencies), which extend to the entire duration of the stimulus (Fujisaki and Kawashima 1970). It appears from their results that the longer the duration of the cues, the stronger the representation in short-term acoustical memory, but that representation may be somewhat weaker in phonetic memory. However, evidence against this dual memory model explanation of CP comes from studies showing that even consonants may not be subject to CP (Massaro 1994). There is evidence from reaction time studies that the time to judge stimuli as being the same is longer when there are within-category differences between consonants compared to when they are exactly the same. Ratings of the degree to which stimuli represent a category also vary when there are within-category differences (Massaro and Cohen 1983).

The original theorists of CP in speech believed CP to be a unique property of speech perception and presumed it to result from the fact that speech is accurately perceived because there is a psychological attachment between speech perception and the articulatory processes that produce speech sounds (Liberman et al. 1967, 1957). Listeners’ implicit knowledge of the articulatory movements necessary to produce a speech sound result in the accurate perception of that sound when heard produced by other speakers. This is known as the motor theory of speech. Only distinct articulatory movements are possible to produce most consonants. There is no articulatory movement that can produce a sound between /ba/ and /da/. When
synthetic speech is produced that varies continuously along the ba/da continuum, according to the motor theory the sound must be perceived as either a ba or da sound but cannot be heard as something in-between the two. The fact that CP has been shown to occur in perception is considered by motor theorists to be evidence in favour of the motor theory of speech perception.

The motor theory of speech suggests that perception of speech is governed by a unique speech system or ‘speech mode’ of processing (Liberman et al. 1967) (see also Studdert-Kennedy and Shankweiler (1970), Studdert-Kennedy et al. (1972) for other evidence of a speech mode), which processes speech differently to non-speech sounds (Liberman et al. 1967). As a result findings relating to speech processes, such as categorical perception, may not occur for non-speech sounds. However, the uniqueness of categorical perception to speech sounds, and the motor theory mechanism as an explanation for CP in speech, has been contested through a number of findings including those showing: that speech perception is more analogue than first claimed (moderated CP); that CP boundary locations can move through context effects; that CP can occur for non-human animals not capable of producing speech; and that CP is observed in non-speech and non-auditory domains.

**Categorical perception in non-speech**

Early CP studies suggested that CP is unique to speech and arises from speech-specific mechanisms and a connection with articulatory processes. However, studies have shown that non-speech sounds are subject to categorical perception. If CP can be found in the processing of non-speech sounds then this may show that a) innate discontinuities in the auditory system are behind CP or b) CP may occur through learning of and parsing into categories of any complex stimuli not just language and c) CP may be relevant to channel compensation when listening to non-speech as well as speech sounds. The following studies provide evidence for CP in non-speech and propose mechanisms behind this CP.

Cutting and Rosner (1974) showed that a sawtooth wave was categorised as a plucked string when the rise time was under 40ms but a bowed string when rise time was over
40 ms. Discrimination performance was also at maximum when the stimuli bridged this category boundary and within-category discrimination was at chance. Miller et al. (1976) has also shown categorical perception for noise/no-noise continua where the time delay between the noise and buzz was altered. The results have been reported to show that categorical perception in speech is ‘merely an auditory process’ (Remez et al. 1980).

Cutting et al. (1976) also showed that like with speech, time between comparisons affects categorical perception in non-speech perception. In a discrimination task, categorical perception occurred for consonants (see Figure 5.6) and a non-significant trend towards categorical perception for vowels was seen. Categorical perception for sawtooth waves (with rise times ranging from 0-80 ms) was found at a 30 ms boundary. Performance for this stimulus was similar to that seen for consonants but CP was only seen with sawtooth stimuli that lasted 750 ms not with stimuli truncated at 250 ms (see Figure 5.7).

![Figure 5.6: Categorical perception for vowels and consonants and the effect of memory. Top row: identification performance (dashed lines) and discrimination performance (solid lines), for consonants (left hand graph) and vowels (right hand graph). Bottom row: discrimination performance as a function of stimulus pair (pairs 3-5 bridge category boundaries, 5-7 and 1-3 are within-category pairs), and of delay between onset of A and onset of X. Taken from (Cutting et al. 1976)](image)

For vowel stimuli it was shown that when the interval between the onset of A and the
onset of X varied between 250, 750, and 1800 ms there was a decrease in discrimination accuracy for within-category stimuli as delay time increased. However, this effect did not occur for between-category vowels and did not occur for consonants (for which within-category discrimination was at chance at all time delays) (see Figure 5.6, bottom row). These results echo those observed by (Pisoni 1973). With the 750 ms sawtooth stimuli there was no decay in discrimination for the between-category comparisons with increased time between stimuli presentations (see Figure 5.6), bottom row. For within-category discrimination accuracy with sawtooth stimuli declined to asymptote within the time gaps used (it appears that asymptote is reached after 250 ms). Discrimination of within-category distinctions was more robust for vowels than sawtooth waves but within-category distinction was possible for sawtooth waves for a short time. Perception is therefore more categorical for sawtooth waves than for vowels, but less categorical for sawtooth waves than for consonants.

This result shows that categorical perception may occur for non-speech sounds and like speech they may be affected by the time gap between comparisons. The results provide evidence for a memory theory of categorical perception, that is affected by the nature and time-course of the acoustical cues. Sawtooth stimuli involve greater reliance on timing information than vowels. Their cues are presented for shorter periods than the vowel’s spectral cues. However, Cutting (1982) notes that longer waves (750 ms) were perceived as more categorical than 250 ms waves. This is contrary to what is expected by the dual memory model. The authors explain this difference by the fact that the short stimuli in this study were not well identified because of inadequate timbral representation.

In conclusion Cutting et al. (1976) claim that both types of stimuli may be assigned to a non-speech-specific long-term memory that stores more abstract features of the sound. This study therefore shows that categorical perception may involve non-speech stimuli being reduced to abstract auditory impressions; what Fujisaki and Kawashima (1969, 1970) refer to as ‘phonemic’ impressions. At the same time as CP, short term timbral memory may be able to store variation where the cues are long lasting rather than transient.

Non-temporal variations within non-speech stimuli were examined for categorical
5.1 LITERATURE REVIEW PART 1: COMPENSATION FOR THE VOCAL TRACT AND PHONETIC CONTEXT

Figure 5.7: CP for sawtooth stimuli and the effect of memory. Top row: identification and discrimination performance, for 750 ms sawtooth waves (left) and 250 ms sawtooth waves (right), as a function of rise time 0-80 ms (lines with stimulus names labelled). Bottom row: discrimination performance as a function of stimulus pair (3-5 bridge category boundaries, 5-7 and 1-3 are within-category pairs), and delay between onset of A and onset of X, for 750 ms and 250 ms stimuli. Taken from Cutting et al. (1976).

Perception by Blechner (1977). The central note in a triadic chord was varied gradually in pitch to create a major/minor chord continuum. Musicians and non-musicians were tested to determine whether each group perceived major and minor chords categorically. Musicians perceived a categorical distinction between major and minor chords, non-musicians did not. Both could perceive differences in the mediant note in the scale when varied along the same pitch continuum but presented alone. Perception for this was not categorical.

CP of chords in musicians suggests an ability to perceive harmonic relations that did not exist for the non-musicians. It also suggests that CP for chords is learnt with experience with these stimuli rather than being innate to everyone. Blechner (1977) compares this ‘categorical perception of the gestalt but not the parts’ to CP in speech; transitions in single formants are not perceived categorically, but phonemes can be. A similar finding has been seen for F3 transitions when presented alone and when presented in a phoneme which varied from /r/ to /l/. The F3 transition is perceived by Japanese and English speakers alike but the /r/ and /l/ distinction is only present for
English speakers (Miyawaki et al. 1975). Blechner suggests that the role of experience in creating categories might apply to any complex sound. The results of this study appear to show, however, that robust categorical storage is less likely to occur for simple non-speech timbres compared to more complex timbres.

### Categorisation

Categorisation theories of speech involve statistical pattern matching to exemplars or prototypes of speech stored in memory (Samuel 1982). The process involves a *sorting* of multiple cues or targets to match to phonemes. There is no warping of perceptual space as is seen in categorical perception (Damper and Harnad 2000). Holt and Lotto (2010) describe categorisation as follows:

> ‘One can conceptualise a segment of the speech signal as a point in space representing values across multiple acoustic dimensions [...] speech stimuli are represented by continuous values, as opposed to binary values of the presence or absence of some feature. Speech perception is the process that maps from this space onto representations of phonemes or linguistic features that subsequently define the phoneme. This is an example of categorisation, in that potentially discriminable sounds are assigned to functionally equivalent classes.’

While there is no warping of perceptual space like with CP and PM but there is a ‘parsing of space’ into categories. Parsing does not appear to result in the same loss of sensitivity to within-category differences but as categories develop with exposure to native language the impact of learnt categories may make within-category discriminations more difficult. This is evident in learning a foreign language. For example Japanese speakers have difficulties distinguishing between /r/ and /l/ in English because they have only access to a single category for these phonemes (Best and Strange 1992). However, categories are not fixed and immersion in a second language can create new categories (Flege et al. 1999, Werker and Tees 1984).

According to categorisation theory acoustic detail remains for most, if not all, speech sounds and is utilised by listeners in pattern matching. The identification of some
productions as better *exemplars* of phonemes than others is thought to be evidence of this discrimination (Kuhl and Iverson 1995). Holt and Lotto (2010) note that a number of recent physiological studies and eye-tracking studies show that detail remains in the speech signal that is accessible during recognition tasks.

The key feature of categorisation is that multiple cues or targets contribute to identification and discrimination of phonemes. It was seen in the discussion of target theories that multiple cues are likely to be relevant to speech perception. No one cue can reliably distinguish between phonemes. The concept of cue trading in speech perception has been extensively researched. For example, ‘if one cue is changed to favour category B, another cue can be modified to favour category A and perception will remain the same’ (Repp 1981). It might be that through this use of multiple cues and redundancy in the speech signal, flexibility is achieved and a combination of relatively few undistorted cues is necessary to speech perception. The most striking examples of this can be seen in the accurate perception of harmonically reduced speech including: *sine-wave* speech where the first 3 formants are tracked for frequency and amplitude over the course of a sentence and replaced with sine waves (Remez et al. 1981) (see Figure 5.8), telephone speech with a bandwidth of 300-3000 Hz, and speech where only information below 800 Hz and above 4000 Hz remains, but is perceived with 90% accuracy (Lippmann 1996).

The mechanisms for speech categorisation are thought to be similar to non-speech categorisation and categorisation in other cognitive domains. An unconscious method of storing past examples and matching new speech to exemplars, involving processing within the striatum is put forward by Maddox et al. (2002). The process has been well researched in learning domains where it has been found that the recognition of objects is good when exemplars of the objects have been stored from previous tasks. Performance with novel stimuli is expected to be worse if such a process occurs (see also Pisoni (1993)). Nygaard and Pisoni (1998) show that practice with hearing a speaker produce words can enhance word recognition for familiar words in noise, but not novel words. This is not explained by experience with the speaker’s vowel space as this occurs in both conditions. This finding supports an implicit exemplar-based theory of speech perception which involves pattern matching to stored exemplars. Specific models of speech categorisation have been adopted from the visual domain, including
Figure 5.8: Taken from Remez et al. (1981). Sine-wave speech. (a) a narrowband spectrogram of the utterance ‘where were you a year ago’, showing harmonic structure as narrow horizontal lines along the frequency scale (b) Wideband spectrogram (c) Narrowband spectrogram of 3 tone sinusoidal replicas. No energy is present except at formant centre frequencies.

classic prototype models (Samuel 1982), implicit decision making models (Maddox et al. 2002), exemplar models (Johnson 1997) and pattern matching models (e.g. Normal a posteriori probability model (Nearey and Hogan 1986)). However, work is needed on these models to fully account for the continuous spectral and temporal nature of speech signals, as opposed to the discrete objects that may be compared side-by-side which are common in visual perception.
Categorical perception and categorisation summary and discussion

Categorical perception and categorisation might be a means of reducing signal variation. The research described in this subsection shows that speech perception can range from categorical to continuous or be somewhere in-between, as the perceptual magnet theory suggests. There is evidence to suggest that consonants are readily categorised—only differences between categories exist for these sounds, when categorised they are entered in to a reduced resolution but robust phonetic memory store. Vowels are not categorically perceived because it is possible to perceive small within-category timbral differences between them, but between-category vowels also appear to enter the robust categorical storage to some extent. The extent to which within-category timbral differences for vowels can be perceived depends on time between comparisons. Experiments suggest that the difference between vowels and consonants is because the short auditory cues that distinguish consonants cannot be stored in the more detailed in short term timbral memory store. Only a more robust but less detailed phonetic memory exists for these cues. For vowels both within-category and between-category differences can be perceived. However, this memory for within-category differences may be short lasting. After time only the perception of between-category differences may remain.

Categorical perception implies a perceived reduction of the variation described in Peterson and Barenys data. CP may provide robust perception in the light of distortion because it prescribes that small variation (not large enough to result in different categorisation) is less likely to be stored. CP appears to occur to reduce variation in any cue that is not part of the stimulus prototype. It may therefore reduce the perceived spectral variation that is caused by loudspeakers and other non-speech transmission channels. CP occurs more readily for timing information. More reliable CP effects have been found for transitions in formants and variation in VOT. This difference may mean that CP may aid compensation for channel variation more for consonants than vowels, and for time varying rather than static stimuli.

CP appears to act by blocking the variation of sounds belonging to the same category and leaving the listener insensitive to this. CP does not position the category boundary or alter category boundary: the category boundary is determined by learning, innate
sensitivity, and/or the prior auditory context. This means that in regions of vowel overlap the listener will still face a problem, as it will not be obvious how to categorise a sound. CP does not appear to be able to address the problem of overlap. Rather, CP appears to imply that once that sound is categorised it is heard it is canonical form, not as a variant of this. For these reasons CP appears to be more useful as a means for combating small spectral distortions caused by the channel rather than fixing the problem of distortions that might cause a problem in identifying a sound because its cues have been pushed from the region of one category into the region of another. Where the channel causes variation that is so large that sounds are categorised differently, CP does not appear to remove this error.

The motor theory of CP is no longer in favour as an explanation of CP as CP is found with non-speech suggesting a general auditory mechanism. However, there is still debate concerning the extent to which CP is caused by innate discontinuities in the hearing system, learning of natural language and other processes including adaptive processing. The dual memory model appears to imply that CP is caused by certain features that can be extracted to a phonetic memory store for a longer term, less detailed representation of the sound while a detailed acoustic shorter-term memory acts for some sounds (e.g. those with longer cues). The extent of storage in categorical form may also depend on the complexity of the stimulus. Single formant transition or pitches are not perceived categorically but the complex sound (the phoneme or chord) may be. Comparisons of the simple attributes in short term memory (if long enough lasting) may be possible but where there are not enough cues to form a category representation, this may not be entered into categorical storage and any discrimination made over longer time periods may be poor (e.g. short term comparisons of pitch are possible but comparisons over longer time periods are poor, except for listeners who possess perfect pitch. There is also the observation that listeners can remember some of the fine acoustical detail of sounds for longer periods than is suggested here. Trained listeners may be particularly adept at storing such representations. It may be possible to extend the detailed storage component of the dual memory model by learning.

Categorisation may occur alongside or instead of CP. This does not imply a reduction in variation due to memory limitations but a use of multiple cues, where some are undistorted, to arrive at the correct labelling of an auditory (or non-auditory) object.
in accordance with a stored exemplar. Like CP this is not speech specific and may aid channel distortion in non-speech perception. However, it is not clear whether, in cases where spectral distortion moves formant frequencies, sufficient remaining cues to phoneme identification would exist to allow for correct categorisation. The within-vowel cues that are not affected by spectral distortion would be few.

5.1.5 Section summary and discussion

This section described part of a literature review to determine potential time-gap sensitive mechanisms of real-world compensation in order fulfil Step 1(a) of the research process set out in Chapter 4, thereby contributing to answering Research Question 1b. The work in this section aimed to provide the information necessary to create a list of mechanisms that may be regarded potential time-gap sensitive mechanisms of real-world compensation if Step 1 Questions A-C can be answered for these mechanisms. It was noted that the literature on compensation mechanisms comes from a variety of studies investigating different aspects of speech and non-speech perception. This section of the review concerned compensation revealed in studies that aim to determine mechanisms of compensation for the spectral effects of a specific channel: the vocal tract; and also studies that aim to determine mechanisms of compensation for the spectral effects of the phonetic context. The compensation mechanisms uncovered in this section are summarised here and the chapter questions are answered. The extent to which these mechanisms may be considered time-gap sensitive and whether Questions A-C can be answered for these mechanisms is discussed in Section 5.4.

The work in this section has revealed a number of mechanisms which, if Questions A-C can be answered, may be potential mechanisms of real-world compensation for loudspeakers and rooms. It has been shown that certain features of speech (certain targets) may result in stable perceptions of speech despite colouration by the VT and phonetic context. There may also be an intrinsic normalisation of distorted targets (compensation using cues within the phoneme) and an extrinsic normalisation (compensation using cues outside of the phoneme). These processes appear to cause timbral constancy in speech perception in spite of spectral variation caused by the VT
and the phonetic context, but the same mechanisms may also aid compensation for other transmission channels such as loudspeakers and rooms.

Chapter questions

Question 5.1 asked: a) To what extent does compensation for the spectral effects of the vocal tract occur in speech perception? b) What are the mechanisms behind this? c) What is the time course of this compensation? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

Question a: can be answered. Differences in VT size and shape between talkers means that the spectral location of formant frequencies (the primary cues to speech sounds) are shifted between talkers. Therefore, despite their high signal to noise ratios, even these cues are not robust to this kind of channel distortion. However, variation does not affect listeners’ ability to correctly identify speech across talkers. Either variation is not large enough to cause important disruption to speech targets, other stable cues exist to ensure correct identification, normalisation/compensation mechanisms exist to normalise distorted cues and compensate for VT colouration, or variation is perceived but is not disruptive due to a categorisation process.

The research has not explicitly quantified the extent of variation caused by different vocal tracts using a mathematical model but it has been shown to be large enough to result in potential disruption. Variation in F1 and F2 between talkers spans a few hundred Hz (Ladefoged and Broadbent 1957, Peterson 1952), and has been shown to result in overlap in F1 and F2 space: F1 and F2 cues for a single speech sound are modulated between talkers to the extent that they may represent F1 and F2 cues for an entirely different vowel when made by different talkers (overlap regions span approximately 100 Hz). Such overlap should result in misidentification of speech sounds. Accurate perception despite overlap indicates the operation of mechanisms that achieve constancy and possibly VT compensation mechanisms. The extent to which constancy occurs is difficult to assess but it seems to be adequate to reduce overlap as accuracy of speech identification is good in single and mixed talker lists.
Some variation between talkers that does not cause overlap is expected to be perceived as this might be necessary to identify the talker. Compensation for VT characteristics is not expected to be complete.

**Question b:** there are a number of mechanisms that may be behind invariant perception across VTs. Phoneme intrinsic mechanisms may be involved. There may be one or a few key spectral targets (e.g. formants) that remain stable despite VT variation, which can be used for identification. Alternatively, there may be a redundancy of cues and a trading of spectral cues for temporal and dynamic cues, in the face of spectral distortion. However, dynamic cues are also affected by VT distortion and temporal information alone has a relatively small role in perception so there is little scope for trading of spectral cues with dynamic or temporal cues. Intrinsic normalisation may occur whereby listeners normalise distorted targets with regard to others that are stable within the same phoneme. F0 and F3 are examples of stable cues that could be used as reference cues for normalisation. This process may move perceived formant targets to a canonical location. However, the intrinsic normalisation methods described appear to primarily adjust for VT length; there are many VT degrees of freedom not accounted for, so perceived F1, F2 locations and other cues relevant to identifying a phoneme still cannot be adequately predicted. Perception of stable target ratios may also result in stability of perception. This is similar to intrinsic normalisation but appears to involve the perception of targets within a phoneme relative to each other, rather than relative to one or two particular reference cues. However, like with intrinsic normalisation using reference cues, unless ratio perception occurs in respect of many ratios, this process is unlikely to adequately describe the VT frequency response and adequately remove VT colouration.

Phoneme extrinsic mechanisms appear to provide further information for compensation for VT characteristics. Prior experience with the speakers’ spectral range or *vowel space* obtained from prior speech may aid perception. Some studies have shown improvement in identification performance with single talker versus mixed talker phoneme lists suggesting that compensation for the VT results from hearing prior speech made by the talker. Ladefoged and Broadbent (1957) describe the mechanism involved as one of relational processing, whereby new speech sounds are perceived relative to the spectral range of the speaker and overlap is reduced. This relational processing explanation
implies a cognitive process is behind this compensation. Some studies have shown no evidence for improvement with prior experience with the same talker with single talker lists and point vowels have been shown not to be useful in aiding compensation. It has been concluded that a large amount of information must be contained in the single phoneme to allow for the good perception of isolated speech sounds and extrinsic compensation is not necessary. However, this conclusion may be based on tests that have not presented listeners with vowels within the region of overlap so the extent that they are testing compensation benefits is not clear. Further, point vowels may not be useful in disambiguating target vowels because insufficient information may reside within these single phonemes to provide a picture of the individual VT. Some researchers have claimed that experience with prior speech is helpful because it provides speech-rate information and normalisation of temporal cues rather than spectral cues.

CP (or the PME variant of this) may also be involved in minimising variation in speech cues caused by VT colouration. CP describes the process of turning continua into categories. Once CP has occurred, robust categorical representations of sounds exist, but much of the acoustical detail is lost, similar to analogue-to-digital conversion at a low resolution. This method of reducing perceived variation does not appear to address the problem of overlap however. Where variation is sufficient to place the target cues in an ambiguous region CP cannot be used to shift cues back to their canonical location. Categorical representations would be formed nonetheless (placing the phoneme in the wrong category). It appears that CP is useful for reducing more minor spectral colouration, but cannot aid perception where cues are moved to regions where a different categorisation would occur. CP occurs more readily for transient cues rather static, so it is not clear whether it can minimise the effects of static spectral colouration or can occur where the sound itself contains static rather than transient cues. CP may also be more likely where the sound is complex and therefore may not reduce channel variation when listening to simple sounds, such as a sine-like tone or noise. Another process, categorisation, may reduce variation. Categorisation theory states that listeners recognise speech tokens regardless of variation because they can use a multitude of cues, which can be matched to an exemplar. This is a process of using redundancy in the signal to arrive at correct object recognition. It is similar to target perception described above, but the listener can rely on many cues rather
than one or two key targets. This method does not remove variation from perception by shifting distorted targets back to their canonical location but the listener may become less predisposed to perceiving variation once categorisation has taken place. However, as with target perception, it might be that all or many of the necessary spectral cues are disrupted by the channel. If speech sounds are primarily determined by these cues, insufficient cues for correct categorisation may exist after colouration by the VT or other channels. Like CP this process does not shift distorted targets back to canonical locations so may be inadequate to remove large distortions which result in shifting targets to a different category region. Both CP and categorisation appear to have limited use in compensation for large distortions that shift target cues into the regions of a different speech sounds. CP and categorisation are descriptive theories. An explanation of the mechanisms behind them is necessary. CP may be a result of learning sound categories or innate discontinuities in the hearing system such as those resulting from limited timbral memory stores which sort time-varying cues differently to longer-lasting static cues and possibly complex stimuli differently to simple. Categorisation is a general perceptual process rather than an auditory process and category processing within the striatum may be involved in this.

**Question c:** the time course of VT compensation mechanisms was given. Perception of the correct vowel using intrinsic mechanisms (target cues, target ratios and intrinsic normalisation) is immediate or almost immediate. Little data was presented stating the time course of the extrinsic VT compensation processes. Prior experience with the VT is necessary for extrinsic compensation. The research describes an active process and Ladefoged and Broadbent’s (1957) study shows that it may require a sentence-length segment of prior speech. However, the spectral range experienced with the precursor, rather than its length, may be important. Compensation for prior spectral range does not occur when there is more than a 10 second gap between the prior sentence and the test sound. CP and categorisation can be considered phoneme intrinsic mechanisms, occurring immediately upon perception of the sound, but there may be a short temporal element to CP.

**Question d:** little information was provided regarding the extent of these mechanisms within a non-speech listening context. Some of the mechanisms may be specific to compensation for the VT and specific to listening to speech, so they may not occur when
listening to other sounds. However, CP and categorisation have been shown to occur with non-speech sounds so these mechanisms may be general auditory compensation mechanisms (see further Section 5.3 and Chapter 6).

**Question e:** the extrinsic mechanism of VT compensation may be considered time-gap sensitive as it is shown that perception relative to the spectral range of prior speech (i.e. shifts in the perception of spectral cues in a test sound in light of the prior spectral context) does not occur when the precursor is separated from the test sound (10s was enough to prevent this compensation in Ladefoged and Broadbent's (1957) study). Intrinsic mechanisms including CP, categorisation, target perception, target ratio perception, and intrinsic normalisation are not time-gap sensitive.

**Question 5.2** asked: a) To what extent is the phonetic context compensated for? b) What are the mechanisms behind this compensation? c) What is the time course? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

**Question a:** variation to a phoneme’s spectrum is caused by adjacent speech sounds, due to co-articulation. Formants move closer to those in adjacent speech, resulting in target formants being located between the target sound and adjacent sound (Lindblom 1963, Lindblom and Studdert-Kennedy 1967, Peterson and Barney 1952). This is known as acoustic undershoot. Such assimilated speech sounds cause a reduction in vowel space and overlap in vowel space, like that caused by VT variation. According to Lindblom and Studdert-Kennedy (1967) overlap spans at most 100Hz. The lack of confusion when sounds are heard next to different adjacent sounds shows that compensation for undershoot may occur. The *overshoot effect* was shown to result in perception of formant frequencies being pushed in the opposite direction to the assimilation caused by undershoot. This pushes perception towards context free targets resulting in compensation for the effects of co-articulation and compensation for overlap. The effect also enhances spectral change between adjacent stimuli.

**Question b:** the physical effect of undershoot is similar to that of speaker variation. In both cases targets are not reached because of limitations of the vocal system and there is movement of the cues in a way that is unique to the talker/prior context.
In both cases mechanisms are required to remove the unique effect of the talker/prior context and shift perception back to canonical targets. The process of compensation for undershoot (overshoot) appears to be similar to the process of extrinsic compensation for the VT described by Ladefoged and Broadbent (1957). In both cases the spectrum of a prior sound (the immediately prior phoneme for overshoot, the whole prior sentence for VT compensation) is sampled and the perception of the target speech cues is in relation to this, shifting the target away from the prior sound in a contrastive direction. Both Ladefoged and Broadbent (1957) and Lindblom and Studdert-Kennedy (1967) suggest relational processing explains these shifts. This implies a cognitive process. However, Lindblom and Studdert-Kennedy (1967) also put forward a peripheral neural adaptation explanation for overshoot. The role of cognitive and neural mechanisms were not tested in the research described, but are discussed further in the remaining sections of this chapter. The intrinsic factors described for VT compensation may also ameliorate the effects of undershoot. Target perception (where any stable targets remain), target ratios, intrinsic normalisation, CP and categorisation may aid perception in light of distortion caused by the phonetic context.

**Question c:** little is known about the time course of overshoot from the research described in this section. It appears that only short phoneme-length sounds are needed to bring about this effect. Approximately 80 ms sounds resulted in overshoot in Lindblom and Kennedy’s study. The recovery of effects when the vowel is separated from its consonantal context, are not described. Due to its similarity with extrinsic compensation for the for VT, overshoot is also likely to not occur when the prior sound is separated from the target sound by silence. Therefore, the process is probably also time-gap sensitive. Based on its short onset time and its possible role in compensation for the immediately adjacent context it is unlikely that the overshoot effect would last longer than the 10 s shown for compensation for the VT. It is likely to have a shorter overall time course and expire quicker than the extrinsic VT compensation process. This is discussed further in the following sections for the literature review.

**Question d:** like compensation for the VT, overshoot may be a specific process that has evolved to remove colouration caused by the phonetic context when listening to speech. It is not clear whether this process can occur to compensate for the effects of other channels when listening to non-speech sounds. Likewise intrinsic methods
such as target perception, that might occur to provide stability in light of undershoot may also be speech specific. However, it is known that CP and categorisation are general auditory processes which might provide channel compensation when listening to non-speech (see further Section 5.3 and Chapter 6).

**Question e:** The overshoot effect is likely to expire with a silent gap between a target vowel and its phonetic context so it is probably time-gap sensitive. This is discussed further in the following sections for the literature review.

**Question 5.3** asked: a) To what extent is the effect of transmission channels other than the vocal tract (e.g. loudspeakers, rooms) compensated for in speech perception? b) What are the mechanisms behind this? c) What is the time course of effects? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

No direct evidence was presented to answer Question 5.3. The tests described in this section involved compensation for speech specific sources of spectral colouration. It may be that the mechanisms described evolved in response to speech specific problems: variation to speech spectra due different VT characterises and variation to speech spectra due to different phonetic contexts and the sluggishness of the VT system. These mechanisms may therefore only been engaged when listening to speech produced by VTs or in varying phonetic contexts. A speech mode of perception may be necessary to engage these mechanisms. It is possible that the mechanisms behind compensation for the VT and phonetic context provide compensation for spectral colouration caused by other transmission channels as no evidence was presented that rules this out. However, evidence in support of this was also not provided here. Further work is needed to confirm this. It was noted that extent of compensation necessary may be different with other channels. Variation to the spectrum of sounds caswed by channels other than the VT and phonetic context may be less extreme so the extent of compensation caused by the same mechanisms may be reduced. Channels such as loudspeakers are more likely to affect wideband spectral cues, which are generally less important in speech perception than formants. However, shifts in spectral cues by other channels, such as loudspeakers, may still occur due to the limits of the those systems and VT/phonetic context compensation mechanisms may contribute to compensation for this.
5.1 LITERATURE REVIEW PART 1: COMPENSATION FOR THE VOCAL TRACT AND PHONETIC CONTEXT

**Question 5.4** asked: a) To what extent does compensation for transmission channels occur when listening non-speech sounds? b) What are mechanisms behind this? c) What is the time course of effects? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time gap sensitive?

The extent of compensation when listening to non-speech sounds was not investigated, as most of the research described involved an entirely speech context. Some perceptual mechanisms that occur when listening to speech may be speech specific (Repp 1982) and a speech mode of perception may be necessary to instigate the compensation described here. Further reviews and/or further experiments are necessary to determine whether the particular mechanisms discussed here are speech specific (see further Section 5.3 and Chapter 6). CP and categorisation have been shown to occur when listening to non-speech sounds, so they might be involved in compensation for any channel when listening to any sound.

It can be concluded that of the mechanisms discussed in this part of the literature review, the mechanism most likely to be a potential mechanism behind the time-gap sensitive component of real-world compensation for loudspeakers and rooms (i.e. the mechanism most likely to be considered time-gap sensitive and provide positive answers to Step 1 Questions A-C) is the extrinsic compensation for the VT mechanism described by Ladefoged and Broadbent (1957). There is sufficient evidence that this is time-gap sensitive and it may have the nature necessary to provide the real-world compensation seen in Experiments 1-3 in Chapter 3 and in Olive et al.’s (1995) work: colouration by the VT and loudspeakers/rooms is similar and both may benefit from prior experience with the channel.

Compensation for colouration by the phonetic context (Overshoot) is another process that may be involved in compensation for loudspeakers and rooms but this situation is less similar to colouration by loudspeakers and rooms. The shorter time course of this effect may be less appropriate for removing the effect of these channels. Both of these effects appear more useful for reducing colouration by the channel than the intrinsic mechanisms. The intrinsic mechanisms appear to rely on redundancy or only few cues from within the same phoneme to determine the nature of the channel distortion and the compensation that must be applied, so may be less
robust. Further, intrinsic mechanisms are not time-gap sensitive. Although they may contribute to compensation for loudspeakers and rooms, they cannot contribute to the time-gap-sensitive component of this.

It is not clear that the same process behind VT/phonetic context compensation is behind compensation for colouration caused by other channels. It may be necessary that the distortion is produced by the VT or the phonetic context or listening to speech may be necessary to engage these mechanisms. Further research is needed to examine this. Further, the effect of the VT and phonetic context on the spectra of speech may be more extreme than the effect of other channels, on speech or other sounds, so these mechanisms may not be applied to the same extent outside of this domain.

The extent to which each mechanism revealed in this literature review can answer Questions A-C and can be considered time-gap sensitive is discussed further in Section 5.4, where a list of potential mechanisms of the time-gap sensitive component of real-world compensation in speech and non-speech listening will be compiled.

5.1.6 Section conclusion

The work in this section contributed to answering Research Question 1b by addressing Chapter Questions 5.1 to 5.4. Answers to Questions 5.1 and 5.2 were given. Prior research showed compensation for colouration by the VT and the phonetic context when listening to speech. Compensation for the VT is not complete but appears to be sufficient to reduce distortion that could otherwise be disruptive to speech perception. Compensation for the phonetic context appears to be sufficient to remove the problem of undershoot and to enhance spectral change between adjacent phonemes. A number of mechanisms appear to be involved in compensation for colouration caused by the VT and the phonetic context. These mechanisms are: target perception, target ratio perception, intrinsic normalisation, categorical perception, categorisation, extrinsic normalisation for the VT and extrinsic normalisation for the phonetic context. Most of these mechanisms are immediate and not time-gap-sensitive but the extrinsic mechanisms have a longer time course. They require time spent listening to onset (sentence length prior listening may be required for VT compensation and
approximately 80 ms for phonetic context compensation). They also offset with silent
time-gaps between prior listening and a test sound (a 10 second gap was sufficient for
recovery of the VT compensation mechanism and it is likely that phonetic context
compensation offsets with a shorter time gap than this). They are therefore time-gap
sensitive.

In the course of addressing Question 5.3 it was established that these mechanisms might
be potential mechanisms of real-world compensation for loudspeakers and rooms but
distortion via the VT and the phonetic context are unique to the workings of the VT
(its limited range and sluggishness) and unique to speech perception. Therefore, these
mechanisms might be unique to compensation for VT effects or unique to listening
to speech. No direct evidence of the working of these mechanisms to compensate
for other transmission channels was provided, and only categorisation and CP were
shown not to be speech specific. Therefore, Question 5.3 and Question 5.4 could also
not be fully answered. Future work should aim to investigate compensation for other
types of transmission channel (i.e. transmission channels generally) and compensation
mechanisms that occur in non-speech listening. This is necessary to determine potential
mechanisms of real-world compensation for other channels, including loudspeakers and
rooms, when listening to other sounds, such as music. A literature review to do this is
provided in the remainder of this chapter (see Section 5.2 and Section 5.3).

Whether the compensation mechanisms described here may be regarded as potential
mechanisms of the time-gap sensitive component of real-world compensation when
listening to speech and non-speech (whether or not Questions A-C can be answered
for these mechanisms and whether they are time-gap sensitive) is discussed further in
Section 5.4. Only when these issues are addressed can a list of potential mechanisms of
the time-gap sensitive component of real-world compensation by complied and Research
Question 1b properly answered.
5.2 Literature review part 2: General compensation for the channel in speech perception

The work in the previous section investigated mechanisms of timbral constancy, including mechanisms of channel compensation. However, this work examined compensation for a speech specific transmission channel, the VT, and a speech specific source of variation, the phonetic context. Compensation for other channels may arise by the same or different mechanisms. Further, the mechanisms behind compensation for the VT and the phonetic context were not adequately described. This section aims to determine mechanisms of timbral constancy and channel compensation in response to colouration caused by those sources and but also to colouration caused by transmission channels more generally. Two auditory effects, the enhancement effect and the spectral compensation effect, are discussed. These effects appear to show specific mechanisms that compensate for spectral variation caused by the phonetic context and the VT and but also appear to show compensation for any other transmission channel when listening to speech. These effects, their magnitude, time course and mechanisms, are described to provide further information on channel compensation mechanisms that may be potential mechanisms of the time-gap sensitive component of real-world compensation for loudspeakers and rooms. Therefore, this section aims to contribute to fulfilling Step 1a(i) of the research process (see Chapter 4). Any of the mechanisms identified here which are time-gap sensitive and for which Step 1 Questions A-C (see Chapter 4 and below) can be answered for speech and non-speech listening can be regarded as potential mechanisms of the time-gap sensitive component of real-world compensation. When one or more such mechanisms are identified Research Question 1b will be answered.

Step 1 Questions:

A) For any compensation mechanism, does the mechanism produce a measurable effect that is indicative of compensation and compatible with providing the real-world compensation seen in Olive et al.’s (1995) experiment and in the
5.2 LITERATURE REVIEW PART 2: GENERAL COMPENSATION FOR THE CHANNEL IN SPEECH PERCEPTION

experiments described in Chapter 3?

B) Is there one or more distinguishing feature of this mechanism (e.g. is it crossaural or does it require a long onset)? This is required to:

- confirm that the mechanism is a unique and distinct mechanism and not a manifestation of another compensation mechanism and

- separate this compensation mechanism from other compensation mechanisms in tests of elimination in step 2.

C) Is the time course of the mechanism such that it is compatible with being an explanation of the real-world compensation seen in Olive et al.’s (1995) experiment and in the experiments described in Chapter 3?

The extent to which the mechanisms discussed in this section of the literature review are time-gap sensitive and the extent to which Step 1 Questions A-C are answered for these mechanisms is mainly addressed in Section 5.4. In order to summarise the information provided in the current section the same chapter questions as described in the previous section of the literature review will be answered:

**Question 5.1:** a) To what extent does compensation for the spectral effects of the vocal tract occur in speech perception? b) What are the mechanisms behind this? c) What is the time course of this compensation? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

**Question 5.2:** a) To what extent is the phonetic context compensated for? b) What are the mechanisms behind this compensation? c) What is the time course? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

**Question 5.3:** a) To what extent is the effect of transmission channels other than the vocal tract (e.g. loudspeakers, rooms) compensated for in speech perception? b) What are the mechanisms behind this? c) What is the time course of effects? d) Can the mechanisms behind this compensation be considered general auditory processes?
e) Can they be considered time-gap sensitive?

**Question 5.4:** a) To what extent does compensation for transmission channels occur when listening non-speech sounds? b) What are mechanisms are behind this? c) What is the time course of effects? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

The enhancement effect is described in Section 5.2.1. The spectral compensation effect is described in Section 5.2.2. The research is summarised and discussed, and the chapter questions answered, in Section 5.2.3. The extent to which Research Question 1b can be answered using the information presented in this section is discussed in the conclusion to this section, in brief (see Section 5.2.4), and in more detail in Section 5.4 of this chapter and in Chapter 7. The answers to Step 1 Questions A-C, and which of these mechanisms can be confirmed to have the potential to contribute to the time-gap sensitive component of real-world compensation for loudspeakers and rooms will be discussed in Section 5.4.

### 5.2.1 The enhancement effect

The enhancement effect may be evidence of a mechanism of compensation for colouration caused by transmission channels, including the VT, and phonetic context. Its nature appears to be particularly suited to compensation for the phonetic context in speech and therefore it may explain the process of overshoot. However, researchers examining the enhancement effect (Summerfield and Assmann 1989, Summerfield et al. 1987) have described it as an effect that might be behind channel compensation generally rather than an effect that is behind overshoot. The enhancement effect describes the phenomenon of enhanced perception of a harmonic component (or number of harmonic components) in a complex tone when that tone is immediately preceded by the same tone with those components missing or reduced (see Figure 5.9 c). Summerfield et al. (1987) state that the reintroduced component ‘stands out perceptually against the background of the existing harmonies’. Viemeister and Bacon (1982) have found that this effect results in increased audibility of the reintroduced component: ‘The audibility of a given target component in a spectral
complex can be considerably increased by exposure to the complex with the target component deleted'. This phenomenon has been known for some time and has also been referred to as ‘negative auditory after-image’ (Cardozo 1967, Schouten 1940, Wilson 1970, Zwicker 1964) because enhancement causes the perception of a spectrum that is the complement of the previously presented spectrum, with peaks where there were previously notches. The enhancement effect has been used to make a target component in a complex tone stand out for easier identification. Simply cueing the component to be picked out, by presenting it alone before the whole complex, did not have a similar effect of making the component easier to discern from the rest of the complex (Cardozo 1967) (Figure 5.9 d). This suggests that the enhancement effect is different to the
effect of simply drawing the listener’s attention to the component and more effective at increasing the salience of individual components within a complex tone.

Enhancement effect tests

Studies specifically investigating the enhancement effect have mostly used non-speech, or synthetic speech sounds (Summerfield and Assmann 1989, Summerfield et al. 1984, 1987). However, there are other studies that also appear to show the enhancement effect that have been conducted with real-speech (these are discussed in the next subsection). The enhancement effect has been found with noise and sine tone stimuli (Viemeister 1980, Viemeister and Bacon 1982, Watkins 1991, Wilson 1970, Zwicker 1964) and appears to be involved in reducing masking thresholds of tones in noise. Simultaneous masking experiments show that masking thresholds of a tone in noise are raised when the target stimulus is introduced later than the masker (Viemeister and Bacon 1982, Zwicker and Fastl 1972). The threshold for detecting a single tone in broadband noise is 10-12 dB lower when the masker is presented continuously, prior to the signal, rather than gated on and off with signal (Zwicker 1964). The enhancement effect has also been found in forward masking experiments. In Viemeister and Bacon (1982) a 2 kHz component in a 200 Hz complex tone was tested for the amount of forward masking that it produced when it followed a precursor consisting of the same tone with the 2 kHz component missing or reduced in amplitude. 8 dB more forward masking was produced when the component in the precursor was initially missing and then introduced compared to when it was presented at the same time as the rest of the complex. The result showed an ‘effective increase in the reintroduced component above its physical level’ and is suggestive of a physiological, rather than a purely perceptual, increase in gain. The enhancement of the masking ability of the component was long lasting. Viemeister and Bacon (1982) found that about 20% of the effect remained after 6.4 s, when the precursor was 2.4 s in length.

Flat spectrum vowel tests

The enhancement effect has also been demonstrated in vowel identification experiments. The Flat Spectrum Vowel (FSV) effect is caused by enhancement (Summerfield and
Table 5.1: An example of the frequencies omitted or reduced from a flat spectrum to create a vowel complement precursor stimulus. Values in Hz. Taken from Summerfield et al. (1987).

<table>
<thead>
<tr>
<th>Vowel</th>
<th>/i/</th>
<th>/a/</th>
<th>/u/</th>
<th>/ɛ/</th>
<th>/ɔ/</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>250</td>
<td>650</td>
<td>250</td>
<td>350</td>
<td>450</td>
</tr>
<tr>
<td>F2</td>
<td>2,250</td>
<td>950</td>
<td>850</td>
<td>750</td>
<td>1,250</td>
</tr>
<tr>
<td>F3</td>
<td>3,050</td>
<td>2,950</td>
<td>1,950</td>
<td>2,850</td>
<td>2,650</td>
</tr>
</tbody>
</table>

Figure 5.10: FSV tests: a flat spectrum precursor sound with three pairs of harmonics representing vowel formants reduced. The harmonics are reintroduced in the flat spectrum test sound. Perception of a vowel with a spectrum that is the inverse of the spectral filtering of the precursor is created in the test sound. Taken from Summerfield and Assmann (1989).

Assmann 1989, Summerfield et al. 1984, 1987). A general description of the method used in FSV studies is given and depicted in Figure 5.10. To create a precursor, a complex tone with a fundamental of 100 Hz and the first 50 harmonic components at uniform intensity is taken (a spectrum with a reduction of 6 dB per octave has also been used (Summerfield et al. 1984)). The frequencies corresponding to formant frequencies a target vowel (see Table 5.1) are then either removed completely, or attenuated by a specified amount (e.g. by 5 dB in Summerfield et al. (1984) and by 5 or 2 dB, in Summerfield et al. (1987)) to create a vowel-complement precursor. The precursor is presented for 1 second. A test stimulus follows the precursor. This consists of a flat spectrum, created by returning the omitted or attenuated harmonics to their previous levels. The test stimulus either follows after a silent gap (1-512 ms, Summerfield et al. (1984), or with a gradual ramp, Summerfield et al. (1984)). The test stimulus is 200 ms in duration (Summerfield and Assmann 1989, Summerfield et al. 1984, 1987). Subjects are asked to identify the vowel heard in the flat spectrum test stimulus and identification accuracy is measured (% correct). The magnitude of the enhancement effect can be measured by the accuracy, above chance, with which a vowel created in this manner can be identified in the flat spectrum test sound.
Summerfield et al. (1984) and Summerfield et al. (1987) have shown accurate identification of the vowel, specified by the complement precursor, in a flat spectrum. The effect was shown to be monaural and did not occur when the precursor and test sound were presented to different ears (Summerfield et al. 1984, 1987). The results show that the enhancement effect enhances perceived increases in amplitude between the precursor and test, so that they are above actual amplitude gain. This can also be explained by saying that listeners are more sensitive to changes in spectrum, than to the level of unchanged components.

Summerfield et al. (1984) also tested whether a vowel complement stimulus that was presented immediately after, instead of before, a flat spectrum test stimulus would induce an FSV. No enhancement was found, showing that the effect does not work backwards. The authors interpreted this finding by noting that the effect enhances onsets of sounds but not offsets. Decremental FSV tests have also been conducted (Summerfield et al. 1987). These tests involved presenting a precursor with a flat spectrum at 46 dB with the harmonics composing a vowel reduced to 41 dB (see Figure 5.11 a, i). A flat test sound at 41 dB followed (see Figure 5.11 a, ii). This process created a flat spectrum vowel effect that was specified by the reduced components in the precursor with 78% correct identification (see Figure 5.11 a, iii). This result occurred because the presentation of the test sound resulted in decremental change for the harmonics not present in the vowel. The harmonics of the vowel remained at the same level. The decrement was enhanced in the perception of the test stimulus so that the harmonics in the test sound that did not specify the vowel were reduced in perceived amplitude compared to the vowel formants. These results show that the enhancement effect exaggerates decreases in amplitude, as well as increases in amplitude, between precursor and test sounds.

**Peaked spectrum vowel tests**

Summerfield et al. (1987) noted that the FSV effect might show compensation for noise, or the spectrum of a transmission channel. The test sound represents a sound source and the precursor represents prior experience listening to the channel. However, they noted that the act of listening to a source with a flat spectrum (the test stimulus), in channels with large spectral notches (the FSV precursor), does not represent real-
Figure 5.11: a) i) Decremental FSV tests: A test stimulus is comprised of formant frequencies at 41 dB and the remainder at 46 dB. (ii) It is followed by a flat spectrum stimulus at 41 dB, (iii) a perception of the vowel is created in the flat spectrum. b) Decremental PSV tests: (ii) A flat spectrum precursor at 46 dB is followed by a test stimulus with vowel formants at 46 dB and the remainder at 41 dB. (iii) The vowel is enhanced.

world listing to sounds through channels well. The authors hypothesised that the benefit brought about by the enhancement effect should arise with more natural stimuli. Summerfield et al. (1987) hypothesised that all that is needed for the FSV effect to occur is prior stimulation and a change in spectral magnitude. It was not necessary that spectral peaks and troughs in a precursor sound were filled. A flat spectrum precursor, thought to be more representative of the milder colouration by a channel (or noise), provided initial stimulation and a vowel shaped test stimulus was created by raising selected formants within a flat spectrum (called a Peaked Spectrum Vowel or PSV). This represented the peaked formants of a normal monophthongal vowel within the same channel (see figure 5.12). Prior presentation of the flat spectrum (the channel/noise) enhanced the identification accuracy of the test sound, compared to a condition without a precursor. Enhancement does not require spectral notches in a precursor to be filled, it can occur wherever there is prior stimulation. The authors confirmed that this is evidence that the enhancement effect may enhance changes in the spectrum of a sound that is heard through a distorting transmission channel or in noise, with a relatively flat spectrum. Enhancement of PSV vowels did not occur where the precursor was presented to the contralateral ear but the same amount of PSV effect was found when a noise precursor was used instead of a flat harmonic complex (Summerfield et al. 1987). The results appear to be caused by the same mechanism at work in the FSV effect and are further evidence that the enhancement effect works to enhance the perception of spectral change relative to static components.
Decremental changes also cause the PSV effect (Summerfield et al. 1987). When a flat spectrum precursor (46 dB) was increased above the level of a PSV test vowel (41 dB with harmonics representing the vowel at 46 dB—see Figure 5.11 b, i and Figure 5.11 b, ii), the decrease in amplitude between the precursor and test vowel was enhanced, creating more amplitude contrast between the background components and the components raised to make a vowel in the test stimulus. PSV stimuli were identified significantly more accurately with a raised flat precursor, than with no precursor (87% versus 76%). Results are compatible with a process that enhances incremental and decremental spectral change.

**Double vowel tests**

Summerfield et al. (1987) and Summerfield and Assmann (1989) provided evidence that enhancement may be involved in improving speech perception in noise and transmission channels. Channels with uneven frequency responses are more common than the flat channels used in PSV tests. These have a distorting effect on the vowel spectrum, which does not occur with a flat spectrum channel (see Figure 5.13 for an illustration of the effect of 2 different channels, with different frequency responses, on a single test sound). A distortion of the vowel formant frequencies can be seen. Enhancement may be particularly important for increasing vowel identification accuracy in this situation. The role of enhancement in compensation for transmission channel spectrum
was tested using a Double Spectrum Vowel (DSV) test. An analogy of listening to vowels through a distorting channel was created by playing a test sound which was a target vowel (a monophthongal vowel) played through a transmission channel with a frequency response equivalent to the spectrum of another context vowel. The precursor was simply the context vowel and represented the same transmission channel. Figure 5.14 b shows this DSV test stimulus alongside the PSV and FSV test stimuli for comparison. The DSV test stimulus was presented with a precursor (i.e. the PSV context vowel—Figure 5.14 a) or was presented alone. It was expected that if amplitude changes are enhanced at the expense of context an extreme version of the enhancement effect would mean that only the target vowel would be heard when the test stimulus was presented. A less extreme version would allow for the perception of the context vowel but at lower intensity than the target vowel. Results showed that identification of the target vowel was possible without enhancement by a precursor, showing that it was possible to separate the mixed vowel test stimulus to determine the target vowel when presented without a precursor. However, identification improved significantly when a PSV representation of the context vowel was presented as a precursor. Subjects reported hearing a sequence of isolated vowels (precursor, then test), rather than a precursor followed by a double vowel. The enhancement effect therefore greatly minimised perceptions of context vowel within the DSV test stimulus. A similar study
Figure 5.14: FSV, PSV and DSV tests. 

(a) Precursors: a 2kHz tone with uniform spectrum made up of 50 harmonics at 41dB is a base. FSV: The frequencies corresponding to target vowel formants are reduced. PSV: harmonics flat. DSV: harmonics corresponding to a context vowel are raised 

(b) Test stimuli: FSV: flat, PSV: target formants raised, DSV: formants corresponding to the vowel in the precursor raised by 5dB and formants of a further target vowel raised further by 5dB.

(c) Perception: subjects perceive an enhanced target vowel.

by Scheffers (1983) showed that perception was ‘like the target coloured by the context of the second vowel’, so it appears that an extreme version of the enhancement effect did not occur and some perception of the channel remains. Presentation of the precursor to the contralateral ear or an orthographic precursor (a card with vowel identity written on it) did not improve performance compared to the no-precursor condition. Additionally, when subjects were asked to identify both vowels within the test stimulus, they could do this above chance. This demonstrated that either they could still hear the context vowel in the test stimulus or they were using the precursor identity to work out the identity of the context vowel.

The time course and magnitude of enhancement

Establishing the magnitude and time course of the enhancement effect is useful for determining the importance of the effect in transmission channel compensation. The magnitude and time course of the effect should be suited to the magnitude and time course of spectral distortion caused by transmission channels. Knowledge about the
extent and timing of the effect is also useful for ascertaining the underlying physiological and cognitive processes behind the effect and for determining whether the effect might explain the time-gap sensitive proportion of compensation seen in real-world experiments. Estimates of magnitude, onset and recovery time of the effect across studies are varied. The description of the magnitude of the enhancement effect is divided into subsections showing magnitude in FSV, PSV and DSV studies.

**Magnitude in FSV studies**

Summerfield et al. (1984) examined how the size of precursor decrements in an FSV test affects vowel identification accuracy. Precursors were created that were 5%, 10%, 15%, 20% and 25% of full vowel complements (the precursors with the formants omitted). The 5% condition represented a vowel complement with smaller formant dips compared to the 25% condition. Figure 5.15 shows that vowel complement precursors that involve a 25% reduction of vowel formants produce 100% identification accuracy of the target vowel in a flat spectrum. Effects begin to occur with small valleys. Summerfield et al. (1984) report that in this experiment a 4.3% complement (2 dB), around a mean of 50 dB, was the minimum valley necessary to create a vowel which can be recognised significantly above chance. Valley depth appears to be related in a linear way to enhancement. The larger the percentage of the full complement (i.e. the deeper the valley), the better FSV identification and therefore the larger the enhancement effect. The results show that the effect has the potential to be large, resulting in identification accuracy of vowels of 100% with small (25%) amplitude contrasts between precursor and test stimulus. To measure the size of the FSV effect more precisely, the degree of actual amplitude increase to formant frequencies in a flat spectrum required to match FSV accuracy can be examined. Summerfield et al. (1987) found that the same amount of amplitude contrast between formants and the rest of the spectrum is needed to perceive a PSV vowel above chance, as is needed when precursor valleys are filled (i.e. FSV vowels are identified with about the same accuracy as isolated PSV vowels). It appears that the enhancement effect is as powerful as actual amplitude gain.

**Magnitude in PSV studies**

The magnitude of enhancement in PSV studies can be compared with the FSV effect
to confirm that the same process is involved in both test types. In Summerfield et al. (1987) enhanced PSV vowels were identified more accurately than FSV vowels created by the same amplitude change. This is expected because both the actual amplitude gain and the enhancement effect are contributing to contrast with PSV stimuli, whereas only enhancement contributes to the identification of FSV stimuli (with a certain level of dB contrast PSV vowels are expected to be identified in isolation with 100% accuracy so ceiling effects are more likely to affect PSV stimuli). The size of the PSV effect can be estimated by comparing the amount of actual amplitude increase to an isolated PSV needed to create the same identification accuracy as an enhanced PSV vowel. An average of 2 dB gain to an isolated vowel is needed to produce the same accuracy as a vowel enhanced by a flat spectrum, at 70% accuracy, averaged across all test vowel levels. The improvement in accuracy for PSV vowel recognition when it is enhanced compared to when not enhanced is similar to the accuracy with which FSV vowels are identified above chance. This shows that enhancement is the same with both methods. Summerfield et al. (1987) tested the increase in identification accuracy brought about by the PSV effect with a number of different increment depths. Subjects were asked

\[ \text{Figure 5.15: The effect of amplitude contrast on FSV identification. Left: percentage of accurate vowel identifications in an FSV test as a function of the \% reduction of formants in the precursor stimulus, compared to full vowel complement (100\% reduction). Right: the effect for decrements between 1 and 10\% in more detail. From Summerfield et al. (1984)} \]
to identify vowels heard in a test stimulus and also to give a confidence rating ranging from 1 to 5 (1 = very uncertain, 5 = very certain). Identification of PSV vowels was significantly improved with presentation of a flat spectrum precursor compared to no precursor (80% accuracy for 2 dB stimuli and 90% for 4 and 5 dB stimuli with enhancement). The gain in accuracy with enhancement reached maximum (about 20% gain) with 2 dB test vowels with no further accuracy conferred by enhancement with larger PSV test stimuli (see Figure 5.16 a). A ceiling effect for PSV recognition may have occurred in this study for vowel test stimuli made with larger increments (3, 4, 5 dB). Such vowels might have left little opportunity for improvement in accuracy with enhancement, as they are well identified in isolation. However, confidence in stimulus identification continued to increase slightly with larger increments between precursor and test. Increases in confidence show that if this ceiling effect was somehow prevented (for example by presenting the PSV test stimuli in noise, thereby making identification

\[\text{Figure 5.16: a) Gain in vowel identification accuracy between enhanced versus non enhanced PSV vowels as a function of the dB contrast between the formants and other harmonics in the test stimulus. b) The difference in confidence between enhanced versus non enhanced condition. Filled circles and open circles show the difference between gated and gradually introduced test stimuli respectively. From Summerfield et al. (1987).}\]
more difficult), further increases in accuracy with enhancement might be observed for 3-5 dB stimuli. This suggests that enhancement with PSVs, like with FSVs, might be proportional to amplitude contrast, with more enhancement with higher contrast (see Figure 5.16).

**Onset time**

Summerfield et al. (1984) tested the enhancement effect, measured by the percentage of correctly identified FSVs, with vowel-complement precursors between 25 and 250 ms in length (divided into 25 ms steps). Figure 5.17 shows that there is an increase in identification accuracy with duration of the precursor. The effect is maximal with a 150 ms or longer precursor. This result suggests that a build up of the effect occurs between 0 and 150 ms and is maximal after this. Viemeister (1980) showed that enhancement affects masking thresholds maximally when a 400 ms or longer precursor is used and that effects begin to occur for precursors of 50-100 ms. The initial onset of the effect is similar to that established by Summerfield et al.’s enhancement effect, but maximum effect in masking appears to require a longer precursor. The onset of the effect was not tested in PSV or DSV studies. It can be assumed that the onset is the same for these stimuli if they represent the same effect as shown in FSV tests.

**Recovery time**

Recovery can be measured by the extent to which the effect persists after offset of the precursor. Summerfield et al. (1984) varied the length of a silent gap between precursor and test stimuli. Gap durations of 1-512 ms were used. Performance was also measured where there was a linear spectral magnitude transition period instead of a silent gap. Figure 5.18 shows the recovery of the enhancement effect (measured by decrease in percentage of vowels correctly identified) as a function of transition period (Figure 5.18 a) and gap time (Figure 5.18 b). With a silent gap performance begins to decline immediately. A 50% reduction in accuracy occurred when a silent gap of over 100 ms separated precursor and test. Complete decay of enhancement in silence occurs, on average, within 500 ms. The transition condition prevented a decrease in identification accuracy for up to 32 ms. This condition meant that the precursor is presented closer to the target stimulus (although with less amplitude contrast, which
Summerfield et al. (1984) suggest is important to observing enhancement). Other studies have reported different recovery times in silence. Viemeister (1980) reports a recovery of 6 s and Wilson (1970) reports recovery at 3.0 s. Estimates of the time taken for the effect to reduce by 37% for a number of studies are listed by Summerfield et al. (1987): 2.0 s, Viemeister (1980); 1.3 s, Cardozo (1967) (Subject1); 1.0 s, Wilson (1970); and 300 ms by both Cardozo (1967) (Subject 2) and Summerfield et al. (1984).

The offset of the effect is approximately 500 ms reported by Summerfield et al. (1984). The onset effect is also quick (approximately 150 ms). This time course is consistent with the time course of the overshoot effect in speech perception, where only compensation for adjacent sounds is necessary and a quick-acting effect achieves this. The time course of enhancement therefore appears well suited to enhancing contrast between phonemes and producing overshoot. However, the varied estimates of the time course mean that, ideally, the suitability of the time course to overshoot should be confirmed. With regards to transmission channel compensation, compensation may benefit from a longer onset time so that the spectrum of the channel can be sampled rather than the spectrum of the most recent sound produced through that
5.2 LITERATURE REVIEW PART 2: GENERAL COMPENSATION FOR THE CHANNEL IN SPEECH PERCEPTION

Figure 5.18: Recovery of the enhancement effect measured by decrease in identification accuracy (% correct), as a function of transition length (top) and silent time gap (bottom). Results from the ‘AH’ vowel should be ignored as it was reported that this vowel was more easily recognised than ‘OO’ or ‘ER’ due to the flat test stimulus sounding ‘AH’ like. Taken from Summerfield et al. (1984).

channel. However, this may not be necessary as the act of enhancing the changes between the stimuli passing through the channel (e.g. the changing speech sounds in a speech source), means that relatively speaking the channel effect, which is static and may be constant for all of the speech sounds heard through the channel, would be diminished. The short recovery time also supports a conclusion that this effect is involved in channel compensation as recovery times for channel compensation as well as overshoot should be short. Too long recovery times could result in inappropriate compensation. For example, when listening to multiple different talkers each with different vocal tract spectra in normal conversation, it is necessary to be able to apply a different compensation when the talker changes. If recovery is not quick the incorrect compensation from one talker might be applied to another. However, a longer time
period over which to sample the transmission channel spectrum, than is seen in these studies, may be advantageous where channels do not change rapidly but sounds heard within them do.

Enhancement mechanisms

The studies on the enhancement effect have discussed the mechanisms responsible for the effect. These may be broadly described as physiological, involving changes in neural responses, or cognitive. It should be noted that enhancement might be a result of cognitive mechanisms alone or neural mechanisms alone but, because these occur at different levels of auditory possessing, both types of process might be involved in creating the effect. The following subsections will summarise the main arguments for a role of the following processes in enhancement: attention, perceptual grouping, categorisation and neural adaptation.

Attention

Attention has been put forward as an explanation of enhancement (Summerfield et al. 1987). For example with the FSV effect, either more attention may be directed to the formants because of their newness when they arrive with the introduction of the test stimulus or attention is directed to the context stimulus because it is cued by the precursor. In the former case this increase in attention to the target might fully explain why the precursor conditions result in better recognition. In the latter case, once the context receives increased attention, it needs to be separated from the target to establish the target vowel. This may be done more easily where the context is cued by the precursor compared to no precursor but this cannot fully explain the enhancement effect. The subtraction/minimisation required to remove the context sound needs to be explained. It is not clear how this would occur considering that, by this sort of cue, attention has been drawn towards the context rather than away from it. It has been shown that presenting either the context or target before the test stimulus is not effective in creating enhancement, when this engages only attentional mechanisms, rather than an auditory mechanism. For example cueing the target by contralateral presentation or orthographic presentation, rather than prior auditory
stimulation did not aid perception of the target (Summerfield and Assmann 1989). These cues should produce attention effects as attention is both cross-modal (Carrasco 2009) and a central process (Broadbent 1985) and therefore can occur when stimuli are presented to different ears. However, these cues did not produce enhancement. It was also seen in Cardozo (1967) and Viemeister (1980) that cueing a component did not aid in picking out that component from the spectrum. This is strong evidence against attention by cueing, as an explanation for the enhancement effect.

Another hypothesis, which is separate from attentional cueing, is that increased attention may occur simply because of the newness of the spectral components that have changed. However, this argument is somewhat circular. Further explanation of how newness results in increased attention is necessary for this to be an explanation of the enhancement effect. Drawing attention to changes by cueing can result in increased attention to those changes, and some changes may be salient enough to draw attention to themselves. The change in spectrum in this case may be large enough to draw attention to itself in the absence of other stimulation and it is possible that this is sufficient to explain why this stimulus is enhanced. To study this in detail research into attention to changing stimuli would be necessary. Both Summerfield et al. and Viemeister and Bacon conclude that increased attention to the target vowel is thought to be an effect of enhancement rather than a cause. Viemeister and Bacon (1982) state that ‘a popped out target might demand attention but we believe this is not why it pops out in first place’. The balance of evidence suggests that other processes are involved in the enhancement effect and it is likely that increased attention to changes between precursor and target is a result of one, or a number of, these processes. An attentional explanation does not appear to explain the monaural nature of the effect.

**Perceptual Grouping**

Perceptual grouping allows for auditory objects to be discriminated by segregation based on temporal, spatial and spectral features (Bregman 1990, Summerfield et al. 1984). One of the most prominent grouping principles is that sounds in different frequency regions that share a common amplitude envelope are likely to have originated from the same mechanical event and are grouped together (Bregman 1990). Perceptual grouping theory offers an explanation for the FSV, PSV and DSV effects. For example,
when a PSV stimulus follows a precursor with a uniform spectrum the harmonics, whose amplitudes are raised to define the formants of the vowel, are grouped and segregated from the components in the precursor. When a PSV stimulus is presented in isolation all harmonics share common amplitude variation and all 50 harmonics are grouped together. Segregation in the presence of a precursor may therefore make the formants easier to perceive (Summerfield et al. 1987). The same is true for FSV stimuli. The reintroduced components are likely to be grouped together because they share a common amplitude envelope. Grouping is competitive and spatial location might be a stronger grouping cue than common amplitude variation (Bregman 1990). When precursor and test are presented to different ears this cue may override grouping by common amplitude envelope and explain why enhancement is not seen with contralaterally presented precursor and test sounds. The effect of grouping alone probably cannot explain enhancement but it may contribute to it, along with other processes. Like attention, grouping allows for the separation of sources from other sounds but some sort of increased attention to the source or subtraction of context would also be necessary to give target vowels within test stimuli extra perceptual weight. For example, grouping may increase separation, which may make it easier for listeners to attend to the relevant stimuli (i.e. target rather than context), which may in itself provide this extra perceptual weight.

Also, grouping may have a role in enhancement but it is not clear how and why grouping happens; ‘to say that grouping explains enhancement begs the question of how grouping comes about’ (Summerfield et al. 1987). It has been suggested that neural adaptation (see below) might be the cause of this grouping. Summerfield et al. (1987) and Darwin (1984) speculate that neural adaptation and perceptual grouping might work together to create enhancement: ‘adaptation might be supplemented by perceptual processes of source segregation. The energy that defines the precursor and the test stimuli start at different times, may be perceived to originate from different sources, and are segregated perceptually [...] the combined result of adaptation and source segregation is that added energy in FSV and PSV stimuli are heard to stand out from unchanging components [...] both perceptual grouping and adaptation are likely to contribute to enhancement’ (Darwin 1984). According to Darwin perceptual grouping is a secondary mechanism which enhances stimuli further after adaptation to the context has occurred.
However, against the argument for an involvement of grouping in the enhancement effect, is the fact that the grouping of the vowel and context is not prevented by a silent gap between the precursor and test. The gap means that all components of the test sound onset at the same time so they should be grouped together (Summerfield et al. 1987). Summerfield et al. (1987) developed a test to examine the grouping hypothesis using PSV test stimuli. A uniform spectrum noise was used as a precursor, instead of a uniform harmonic spectrum. This noise had approximately the same excitation pattern computed to formulas of Moore and Glasberg (1983). The same amount of enhancement of PSVs was observed with this precursor as with a harmonic test stimulus. Grouping theories predict that the noise is grouped separately from all 50 harmonic components. Further, increased forward masking (Viemeister and Bacon 1982) would also not occur if grouping explained enhancement as grouping alone cannot increase masking. Summerfield et al. (1987) conclude that ‘perceptual grouping should probably be seen as the beneficiary rather than the cause of enhancement’.

Categorisation

Summerfield and Assmann (1989) discussed categorisation or Categorisation-Guided Separation (CGS) as a mechanism contributing to the enhancement effect, in particular, for DSV identification. The CGS explanation of enhancement with a precursor is that stored knowledge of the spectrum of the vowel that is exemplified by the precursor is used to identify the spectrum to be subtracted to diminish the effects of context (Zwicker 1984). It is shown that CGS might occur in the perception of isolated DSV stimuli. Listeners can identify both constituents of a double vowel significantly more accurately than chance, without a precursor (Scheffers 1983, Summerfield and Assmann 1989, Zwicker 1984). Zwicker (1984) suggests that the auditory spectrum of the double vowel is searched for a spectrum that best represents a single vowel or ‘one of the stored templates of vowel identities’. This is then subtracted from the spectrum, to reveal the second vowel. The test using orthographic and contralateral representations of the context vowel to enhance performance of DSV recognition can also be used to examine CGS (Summerfield and Assmann 1989). All three precursors—auditory, orthographic and contralateral—should enhance performance if CGS is used, because all precursors indicate the template of one of the vowels in the DSV stimulus. The fact that no enhancement of target identification occurred with orthographic or contralateral
precursors is evidence against the use of knowledge of stored exemplars, to separate a DSV stimulus and enhance the target vowel. Haggard (1973) found a similar result: listeners were asked to identify words whose spectral components had been transposed to higher or lower frequencies randomly from trial to trial. An introduction phase processed with the same distortion helped identification but prior information as to the nature of the distortion did not. This suggests that listeners cannot use mere knowledge of distortions to discount their effects. These results fit with everyday experience. Humans cannot perceptually adjust for a distortion simply by knowing the nature of that distortion.

It is apparent that knowing the distortion is not enough to remove the distortion. A method of subtraction or minimisation before perception is required to fully remove the effects of the channel. CGS as an explanation of enhancement, rather than isolated DSV separation, appears to be inadequate. CGS does not appear to explain how, once the appropriate exemplar has been determined, the effect of the remaining context-vowel spectrum is minimised in perception. While CGS may be involved in separation where there is no precursor it does not seem to explain increased vowel recognition with a precursor by itself.

**Neural Adaptation**

Early studies on the enhancement effect assumed that adaptation in the peripheral hearing system was behind the effect (Cardozo 1967, Wilson 1970). ‘Simple neural adaptation’ (adaptation of neural firing rate) can explain the enhancement of new sounds relative to context sounds. This process is thought to involve a decreased neural response with continuous stimulation in frequency sensitive channels. Such adaptation has been seen previously in auditory nerve fibres (Kiang et al. 1965, Smith 1979, Smith and Zwisiocki 1971). Summerfield et al. (1987) set out 3 stages to show how adaptation might cause enhancement:

1. The auditory response of components making up the precursor declines over its duration because of adaptation, while the response produced by newly arriving energy that is added to existing components is time invariant.

2. When components that define the test stimulus are raised in level a time invariant
increment is added. These increments produce a greater auditory response than the continuous components. The ratio of new energy to existing energy is larger; this allows new energy to stand out perceptually, as illustrated by Smith (1979) in Figure 5.19.

3. The resulting difference may be further increased by adaptation of suppression (see below).

When an increment occurs to a flat spectrum there should be a perception of greater level because the ratio of new energy to existing energy is greater than when increments occur at the same time as the rest of the harmonics (as in an isolated PSV). Thus adaptation accounts for FSV, PSV and DSV enhancement. The model can also explain decremental enhancement. A decrement is also time invariant so the proportion of decremental change when imposed on the adapted stimulus will be larger than when imposed on an unadapted stimulus. The model proposed by Smith (1979) gives an indication as to the ease with which an increment or decrement can be detected. The later an increment/decrement to the pedestal (the precursor) occurs the easier it will be to detect, until the continuously presented stimuli are fully adapted. This has been confirmed by research that shows that performance in vowel identification increases when the precursor is presented for longer periods up to approximately 150 ms (Summerfield et al. 1984) at which point the precursor might have fully adapted.

The finding of no enhancement effect with crossaural presentation is evidence that the enhancement effect is peripheral in origin (Summerfield et al. 1984, 1987). This supports the case for neural adaptation in the auditory periphery: cognitive processes involve stimuli at both ears. However, evidence to support a simple peripheral adaptation account is mixed. Against a conclusion that enhancement is caused by simple peripheral adaptation are findings that the development and recovery of the enhancement effect may be longer than those shown in primary nerve fibres in human and animal studies. The onset of the effect represents the length of precursor stimulation necessary to reach maximum adaptation and maximum enhancement effect. If adaptation is involved the time taken to reach maximum enhancement should match the time taken to reach maximum adaptation. Neural adaptation in the auditory nerve measured in the guinea pig shows a high spike rate at onset of a tone (800
Figure 5.19: Enhancement and ‘simple’ neural adaptation: amplitude increments are time invariant but continuous ‘pedestal’ tones decline in discharge rate over time because of adaptation. An intermittent tone will produce the same increment in discharge rate whenever it is presented but with progressive presentations the pedestal will produce lower discharge. The ratio of increment to pedestal discharge will increase in favour of the increment with time. $R_i/(R_i + R_{p(i)})$ will become larger (Smith 1979).

spikes/s for 1-2 ms, which falls rapidly within 5-10 ms and reaches asymptote at a rate of approximately 150 spikes per second (Yates and Robertson 1980). Adaptation of discharge rate in individual primary auditory nerve fibres and adaptation of the compound action potential in the rat is reached within about 50-60 ms (Westerman 1985). In general the results show a similar time to maximum adaptation as is shown in the onset of the enhancement effect (approximately 150 ms) (Cardozo 1967). Summerfield et al. (1984) state that ‘there is no evidence that cognitive processes have extended the effect beyond the temporal limits expected from its adaptation component’. However, a study by Viemeister (1980) found an onset time of 800 ms in a masking experiment. This was declared to be unlike maximum adaptation time in the peripheral auditory nerve.

Recovery times of enhancement in silence should match recovery of neural adaptation. Recovery from adaptation of the compound action potential in human listeners is 380 ms (Eggermont 1972). Recovery of adaptation in the auditory nerve of the guinea pig is complete in 150 ms (Yates and Robertson 1980). Several reports of recovery times for enhancement are longer than expected for simple adaptation. Wilson (1970) reported a 3 s recovery time and Viemeister (1980) reported 6 s. Shorter reported recovery times do support a peripheral simple adaptation explanation. Summerfield
et al. (1984) showed recovery of enhancement of approximately 500 ms and Cardozo (1967) reports recovery within 300 ms. There have been few studies investigating the physiological processes in enhancement directly. However, one such physiological study has found evidence for simple adaptation in auditory nerve fibres in guinea pigs during the enhancement process (Palmer et al. 1995). However, Holt and Rhode (2000) found no evidence of adaptation in the auditory periphery during enhancement, but note that neural adaptation may occur at higher centres.

**Adaptation of suppression**

Summerfield et al. (1984) and Summerfield et al. (1987) explain how stimuli may be enhanced by a relative reduction in the perceived level of the background stimulus through simple adaptation (reduction in firing rate). However, this explanation does not account for the finding in Viemeister and Bacon (1982) of an ‘effective amplification of the level of a reintroduced harmonics’, which results in increased power to mask another component. Simultaneous-masking experiments show a decrease in the ability of continuous precursor components to mask reintroduced components (Viemeister 1980, Zwicker and Fastl 1972) but, more unusually, forward masking experiments (Viemeister and Bacon 1982) have found that 8 dB more masking was produced by a 2 kHz component that was enhanced by omitting it from a tone complex and reintroducing it. It is not just that the pedestal (the channel) is reduced but the signal is also enhanced physically as well as relatively. Viemeister and Bacon (1982) confirmed the amplification of the reintroduced component by presenting a precursor stimulus in the left ear, followed by a test stimulus to the left ear and a comparison sound to the right ear. Subjects were required to adjust the levels of components in the comparison to match the perception of a vowel created in the test stimulus. Results showed both a decrease in the level of existing components and an amplification of the level of the reintroduced harmonics. Summerfield et al. (1987) replicated this finding. These results suggest a role for adaptation and adaptation of suppression in the enhancement effect (Viemeister 1982, Sidwell 1985). The new component is thought to have increased excitation because the ability of the continuing components to suppress nearby energy is adapting, as well as the firing rate of the continuous components simply decreasing. This allows new components to achieve greater levels than they can without a precursor (Houtgast 1972).
Byrne et al. (2011) offers more recent support for the involvement of adaptation of suppression in the enhancement effect, showing that a target is perceived to be more intense after presentation of a precursor using an inter-aural level centring task. The role of simple adaptation was not tested so it is not clear if this occurred alongside adaptation of suppression in this experiment. However, Palmer et al. (1995) found no evidence of increased firing rates of newly introduced components using an FSV enhancement paradigm in the auditory periphery of the guinea pig, which would be expected if adaptation of suppression occurs (although simple adaptation of the continuous components was found).

Because of mixed behavioural and physiological results for peripheral adaptation, research has begun into neural responses at central sites. These studies have found adapting inhibitory responses in the inferior colliculus of marmoset monkeys during enhancement (Nelson and Young 2010) and adaptation in the primary auditory cortex (Franosch et al. 2003, Norena 2003) during enhancement-like phenomena. The peripheral origins of the effect are now in doubt but lack of interaural transfer must be explained if central processes are involved. Efferent mechanisms may be involved such as the efferent medial olivocochlear pathway, which occurs centrally but acts on the peripheral response at each ear, creating ipsilateral and contralateral dynamic range adaptation effects. Overall, the peripheral adaptation argument is strong with both simple adaptation and adaptation of suppression causing effects of a similar nature to enhancement and with a similar time course, but further investigation of these mechanisms is required.

**The enhancement effect summary and discussion**

The studies described in this section show that changes in the form of increments or decrements in level between a precursor and a test stimulus are enhanced perceptually. This has been said to improve the identification of changing spectra (i.e. speech) when played through distorting transmission channels and reduce the channel perception relative to the changing sound source (Summerfield et al. 1984, 1987). The effect may also be behind the perceptual overshoot effect in speech perception as the spectra not present in a prior phoneme are perceptually enhanced in a following phoneme,
resulting in compensation for the spectrum of the prior phoneme and contrastive shifts in categorisation between two adjacent vowels. It was also noted by Summerfield et al. that this effect may improve the perception of speech in background noise. Compensation for the effects of noise on listening is not of primary interest to this thesis but the process of compensating for noise and compensating for spectral distortion caused by a static channel is likely to be similar. Work by researchers in the field of perceptual noise-reduction may be able to confirm the role of the enhancement effect in compensating for noise as well as channel colouration and this field of research could be explored for further mechanisms that may be involved in compensation for colouration caused by transmission channels.

The enhancement effect is not crossaural suggesting a peripheral auditory cause or other monaural mechanism. The enhancement effect appears to only enhance onsets as it was not shown to work retrospectively. This fits with the peripheral adaptation explanation which also enhances onsets and does not work backwards. It is not clear whether simple adaptation, adaptation of suppression, or both, explain enhancement. Also the time-course of the effect does not exactly match that of known peripheral adaptation responses. Some studies have not found physiological evidence of peripheral neural adaptation causing enhancement and more recently neural adaptation at central sites has been examined. Other, cognitive, explanations of enhancement were put forward but it was concluded by the authors that these are more likely to be beneficiaries of the effect, or contributors to the effect, not primary causes. However, the perceptual grouping explanation, in particular, appears to remain a candidate explanation.

It is difficult to quantify the magnitude of the enhancement effect in terms of an increase in accuracy per % amplitude change, or in decibels. This is because the effect varies with the length of precursor, the time gap between precursor and test stimuli, and the nature of the stimuli. A comprehensive set of conditions is yet to be tested in a single experiment. The magnitude of enhancement appears to be positively related to amplitude contrast, with more enhancement occurring with greater amplitude contrast between precursor and test (Summerfield et al. 1984). Measured in decibels it appears that overall enhancement results in a few dB relative increase in the signal compared to the the channel (of the order of 20% increase in vowel identification accuracy) which may be important in enhancing spectral differences between speech sounds, which are
sensitive to spectral distortion. This level of enhancement may not offer a notable reduction in channel effect when listening to other sounds if they are less sensitive to colouration by the channel. There appears to be some discrepancy between the amount of enhancement of vowel spectra reported in FSV and PSV tests and the reported reduction of the channel in DSV tests, where it has been reported that the channel is reduced almost completely in one test. It appears that on balance the evidence suggests that any spectral change is enhanced but the unchanging spectrum of the channel is still perceived at a lower level compared to the stimuli changing within the channel. Everyday listening tells us that channel effects are still perceived in most situations but do not dominate perception so compensation via this mechanism is not expected to be complete. The exact magnitude of enhancement both in terms of increased perception of the vowel and decreased perception of the channel needs further investigation before the extent to which the enhancement effect can be said to compensate for transmission channels or overshoot can be determined.

Results relating to onset and recovery of the enhancement effect show that maximum enhancement is created with a precursor of 150 ms but onset time of up 800 ms has been reported in a masking study. These onset times indicate that the enhancement effect is mainly dependent on the spectrum of the precursor immediately before the test and up to 150 ms prior. This short onset appears to result in an effect that has a time course suited to explain overshoot. The effect will enhance the spectrum of a phoneme relative to the immediately prior phoneme thereby compensating for the spectrum of this phoneme. Speech spectra change rapidly and this time course appears appropriate to compensation for the effect of the prior phoneme but not others, as is necessary for overshoot. For transmission channels, which remain unchanged for longer a longer onset period may be appropriate as there may be benefits of sampling the channel for longer to get a better picture of the spectrum of the channel. However, it should be noted that simply by increasing spectral change in the sound source, the channel can become relatively diminished. This enhancement of spectral change relative to the channel appears to be the means by which Summerfield et al envisage channel compensation via the enhancement effect. Therefore, the enhancement effect may be said to cause channel compensation as well as overshoot. Further, the quick onset process of enhancement might also be desirable for channel compensation.
changes occur between channels, for example when multiple speakers are taking it in
turns to speak, short onsets mean that a change in compensation with each new vocal
tract can be applied quickly. The onset of the enhancement effect may be compatible
with perceptual overshoot and transmission channel compensation. However, in terms
of its role in channel compensation more work is needed. It is generally not clear
whether the onset is adequate to offer channel compensation as longer listening to a
static channel is likely to be beneficial.

Recovery times of approximately 500 ms on average occur. These also appear to be
of a short enough duration to work to enhance spectral contrast between phonemes.
This recovery time may also be useful in channel compensation. Gaps in running
speech are not usually longer than this so this recovery time may ensure that the same
compensation is applied, without significant recovery with gaps between words, but
also that quick changes in the compensation applied can be made where there is a
change in transmission channel. However, a longer recovery time may be advantageous
for compensating for static channel colouration.

Firm conclusions about the magnitude and time course of the enhancement effect are
prevented by variation across studies. It is acknowledged that knowledge about the
time course and magnitude is necessary to determine whether the effect is sufficient to
account for perceptual overshoot and/or transmission channel compensation. Different
measurements are due to the use of different stimuli heard at different levels and in
different test paradigms. A comprehensive study should therefore be conducted to
examine the time course and magnitude of effects with comparable stimuli. It is not
clear if the effect is large enough to fully explain constancy of speech across different
talkers’ VTs or across other transmission channels. The time course is more well
established but further work would be beneficial to confirm the short-time course of
the effect. It can be concluded that the effect is time-gap sensitive as compensation
for the prior sound is not applied when there is a gap of over 500 ms between this and
the test sound.

The extent to which the enhancement studies show a general auditory process is not
entirely clear. Masking studies clearly show a non-speech effect and the apparent
peripheral adaptation nature of effect suggests a non-speech specific mechanism but
Summerfield et al.’s experiments have elicited speech perceptions during the test and speech specific mechanisms can therefore not be ruled out. A speech mode of perception may have been engaged. It should be further determined that the results of Summerfield’s tests show a non-speech process. Masking tests firmly show enhancement occurring in tests that do not involve speech but, the sort of enhancement effect seen in Summerfield’s tests—which more clearly shows an effect that may cause channel compensation—may be different be from the enhancement effect seen in masking tests. Summerfield’s work uses stimuli with rich broadband spectra like real vowels filtered by broadband channels whereas masking tests measure enhancement effects with sine tones and narrow noise bands. Effects like two tone suppression, which appear to be different to enhancement, may be behind the enhancement seen with the simpler stimuli. Therefore, the enhancement tests (e.g. FSV/PVS/DSV tests) should be conducted with non-speech stimuli with rich spectra. It is also observed that Summerfield’s studies used simplified speech stimuli (e.g. the vowels lack spectral transitions and vowel inherent spectral change). It is possible that because these stimuli lack many of the features of real speech, the enhancement effect seen here does not represent an effect which will occur when listening to real speech. It may be that this is not a mechanism of channel compensation that works with real-speech. However, enhancement studies with real-speech are described in the next section and appear to show enhancement with real-speech.

### 5.2.2 The spectral compensation effect

Summerfield et al.’s studies appear to show a channel compensation effect with basic synthetic stimuli that elicit speech perceptions. It may be that the effect is different with real speech stimuli. A series of studies by Watkins and colleagues (Watkins 1988, 1991, 1999, Watkins and Makin 1994, 1996a,b) have extended the research of Summerfield et al. to examine enhancement further. These studies have broadly adopted the method of Summerfield et al. but have introduced differences to explore the nature and mechanisms of enhancement, including the extent to which it is speech specific. To examine this real-speech sentences were used as precursors. In the process of examining the enhancement effect with these stimuli another similar ‘spectral
compensation’ effect has been elicited that may represent an additional mechanism of compensation for the channel that occurs with real-speech stimuli. Findings relating to the enhancement effect and the spectral compensation effect roles in transmission channel compensation are described here.

Watkins (1991)

![Diagram](image-url)

**Figure 5.20**: Watkins (1991): probability of a ‘Itch’ response with (Top) an unfiltered male carrier (precursor) as a function of the interpolation of the test sound from ‘Itch’ to ‘Etch’ (x-axis). Bottom: the shift after a precursor filtered for inverse of E (circles) or I (Squares). The boundary (the 50% response point) is moved towards the vowel whose inverse has not been applied via filtering. Taken from Watkins (1991).

In a series of experiments by Watkins (1991) real-speech sentences (spoken by the same
5.2 LITERATURE REVIEW PART 2: GENERAL COMPENSATION FOR THE CHANNEL IN SPEECH PERCEPTION

speaker as the test stimulus) or noise was used to form precursor stimuli. These were filtered by an *inverse-vowel shaped* channel that aimed to represent colouration caused by a typical channel as in Summerfield’s DSV tests. More realistic ‘full-spectrum’ vowel filters were used, unlike the notch/band pass filtering used previously. As in Summerfield’s work this created *vowel-complement* precursors. This filtering was expected to create the perception of vowels in test sounds, where enhancement occurs. In an all speech test a precursor sentence ‘the next word is’ was spoken by a male voice and was filtered (or presented unfiltered). The filtering applied was either the complement of a natural /e/ vowel or the complement of an /i/ vowel. After listening to precursor sentences listeners were required to identify unfiltered test words. A continuum of test vowels was created that had the spectral shape of /i/ /e/ or an interpolation between these. These were used to form a test word continuum from ‘Itch’ to ‘Etch’. Listeners heard test words from across the continuum presented in random order after unfiltered or filtered precursors. An equal number of test stimuli from either side of the Itch/Etch boundary were presented. An increased tendency to hear ‘Etch’ in the perception of the continuum resulted when the precursor was filtered with the inverse of /e/ and a tendency to perceive ‘Itch’ occurred when the precursor was filtered for inverse /i/. This moved the categorisation boundary (the point on the test sound continuum where equal numbers of /i/ and /e/ are heard) towards /i/ and /e/, respectively. When the test stimuli were presented with an unfiltered precursor the identification boundary between perceived ‘Itch’ and ‘Etch’ responses lay in the middle of the continuum (equal numbers of ‘Itch’ and ‘Etch’ were perceived), see Figure 5.20.

The categorisation boundary shifts in the opposite direction to the precursor spectrum indicate perceptual enhancement of the frequency regions in the test vowel that are reduced by filtering to the precursor. As in Summerfield’s studies the results suggest enhanced sensitivity to spectral magnitude change compared to unchanging spectra and appear to show the same enhancement effect. The results were also interpreted by the authors to show a restoration of perception of a filtered sound to that of an unfiltered context; they noted that this is supported by the fact that parallel experiments show that the presentation of a ‘distorted’ test sound (e.g. a sound filtered to be an /e/ continuum endpoint stimulus) is perceived to be ‘undistorted’ (e.g. a mid point on the /i/-/e/ continuum) when the precursor is also filtered for the same distortion (Darwin...
et al. 1989, Haggard 1974)). With the use of speech test stimuli and the categorisation boundary shift measure in Watkins’ study where formant adjusted precursor sentences caused the same contrastive effect on speech categorisation, similarities can be observed with the effects seen in Ladefoged and Broadbent’s (1957) work.

While the results are similar in nature to the enhancement effect and also the results of Ladefoged and Broadbent (1957), Watkins (1991) found an effect larger in magnitude than those seen in Summerfield et al’s work (and his own similar enhancement study (Watkins 1988)). Watkins hypothesised that the larger effect seen in the current study indicated an effect other than the enhancement effect was present and that this effect may only occur when sentence-length natural speech precursors are used; not with the shorter synthesised-speech or noise precursors in Summerfield’s work. If this hypothesis is correct, Watkins’ (1991) result may show an effect more similar to Ladefoged and Broadbent’s (1957) than Summerfield’s, as that study also used sentence length speech precursors. To test whether the effect observed in Watkins’ (1991) study is different from enhancement, a number of sub-experiments were conducted.

Tests with different ears and directions

The enhancement effect was eliminated through presentation to different ears (Summerfield et al. 1984). As discussed in the previous subsection, this is likely to show a monaural effect that is a result of short-term neural adaptation at peripheral sites (though it may be due to other monaural processes that originate centrally). If the boundary shifts in Watkins’ (1991) work are also due to peripheral adaptation, the shift would be expected to be eliminated with presentation of a precursor and test sound to different ears. In a crossaural test a reduction in boundary shift did occur with contralateral presentation, but the boundary shift remained significant, see Figure 5.21 left panel. The reduction in boundary shift indicates a role for a monaural mechanism. However, the fact that some boundary shift occurs when the presentation of the precursor is at a different ear to the test sound indicates a central process is also behind the boundary shift. This was not seen in Summerfield’s studies and suggests that two separate processes are at work in Watkins’ study. This central effect appears to be specific to the sentence-length speech precursors or the speech stimuli used in Watkins’ study and did not occur with Summerfield’s slightly shorter static precursors.
Watkins observed that when the precursor and the test sound are presented to different ears the situation is similar to when sounds arrive from different directions. It was hypothesised that the component of the compensation effect that appears to be monaural may not be monaural, but central and direction sensitive, breaking with direction change (as this suggests the channel has changed). The effect was therefore tested to determine whether it reduced when the test and precursor sounds were presented to both ears (resulting in equivalent stimulation to the monaural condition on the basilar membrane at each ear), but from different directions (causing inter-aural time differences).

A boundary shift occurred in this case, but was reduced compared to the condition where the test and carrier were presented from the same direction, see Figure 5.21 right panel. Because the boundary shift is reduced compared to a same direction condition
this shift must reduce with direction change. The stimulation was equivalent at both ears so this reduction is not accounted for by the absence of the monaural component. Any reduction with direction change where the monaural stimulation is equivalent must be due to a central direction sensitive process (as a monaural process cannot be direction sensitive). This central direction change reduction must contribute to the reduction of the effect with crossaural listening, showing that in the crossaural test that component that reduces with crossaural presentation is not solely explained by a monaural mechanism. As with the crossaural test some shift remained with direction change. Because this remained with both tests some of this remainder must be central in origin. However, there is less reduction in shift with the direction change condition compared to the crossaural condition and this represents a small contribution from a monaural effect, which is likely to be caused by Summerfield’s enhancement effect. Therefore based on the results of the crossaural and direction sensitivity tests it can be concluded that most of the shift with speech sentence-length precursors is due to a central compensation effect that is partly direction sensitive. A monaural effect is present but appears small.

This central effect must be caused by a different compensation mechanism to the enhancement effect, which is not crossaural. Therefore, it is distinguished from the enhancement effect and labelled the ‘spectral compensation’ effect. Watkins concluded that the spectral compensation effect in this study may be due to the use of a speech precursor or at least a precursor that contained spectro-temporal variation unlike Summerfield’s static noise and static harmonic complex precursors. The longer length of this stimulus may also have played a role in bringing about this central effect. Watkins observed that the spectral compensation effect appears to provide compensation for the Long-Term-Average Spectrum (LTAS) of the prior context. This was because both the enhancement effect and spectral compensation cause a boundary shift in the opposite direction to the filtering to the precursor, but with the spectral compensation effect the compensation applied appears to be sampled over the course of the sentence, rather than simply the immediately adjacent sound. However, no specific test was done to confirm this.

Tests with noise stimuli
Watkins observed that the larger spectral compensation effect may only be present with speech precursors: ‘speech engages a more effective means of compensation’. To test the specificity of the spectral compensation effect to speech an experiment using speech shaped noise precursors was conducted. Noise precursors had the same LTAS as speech and the temporal envelope of speech but lacked spectro-temporal variation. In this respect, these precursors are more similar to the static spectrum precursors used in Summerfield’s studies. Watkins hypothesised that, because the speech stimuli and the noise stimuli contain the same LTAS, a similar size boundary shift would be expected if the effect is not speech specific. If the spectral compensation effect is speech specific then speech processing would not be engaged in response to noise precursors and the boundary shift would be reduced. Any boundary shift present with noise is likely to be a result of a general auditory process that applies to any sound.

This test showed a significant but small boundary shift with noise that was smaller than that observed with speech precursors. Watkins concluded again that this shift represented a peripheral component to the boundary shift and that this is a general auditory component. This component appears to represent Summerfield’s enhancement effect. The absence of the larger spectral compensation effect with noise indicates a role for speech specific processes in the spectral compensation effect, or at least processes that require spectro-temporal variation in the precursor that is not present in the noise precursors.

Rather than its size, the defining feature of the spectral compensation effect is its crossaural nature. Watkins tested the noise precursor to confirm that a crossaural effect was not elicited. A crossaural test was conducted, as with speech precursors, where the noise precursor was presented to one ear and the test sound to the other. This manipulation reduced the boundary shift to non-significance. In a direction change test there was no influence of direction change. The effect was also notably increased with the removal of the 160 ms gap between precursor and test sounds presented in Watkins previous test. This showed that the effect had recovered substantially with a 160 ms gap between precursor and test. These results would be expected if the shift with noise exhibits only the monaural enhancement effect and the mechanism has a peripheral adaptation origin.
Results with noise precursors appear to add further support to the hypothesis that compensation for the transmission channel occurs by two mechanisms: a monaural, non-direction-sensitive peripheral adaptation mechanism and a binaural central mechanism that is partly direction sensitive. The monaural effect seen in these studies appears to be a manifestation of the enhancement effect. The central process is new and is termed the spectral compensation effect. The central process appears to be specific to when speech precursors are used, but the speech in the test sounds may also have a role in instigating this effect. Further, the effect may not be specific to speech. It may simply require spectral variation in the precursors/test sounds, which was not present in Summerfield’s test. Where speech precursors are used the monaural peripheral effect appears but is small. The size of this effect may be in-part due to the 160 ms gap in the speech precursor tests, but it may also show that when the stimuli are not static, like those used in Summerfield et al’s work, the enhancement effect is less strong.

**Testing the effect of reversed speech carriers**

It was observed that the spectral compensation effect may be speech specific. Reversed male and female speech precursors were used to test precursors that are unlike natural speech. Both precursors gave a substantial boundary shift. Watkins concluded that the spectral compensation effect occurs when the carrier is unlike normal speech. This supports a hypothesis that the spectral compensation effect is not speech specific but occurs whenever there is spectro-temporal variation in the sound. If it is only dependent on spectral variation, the spectral compensation effect may work wherever there is spectro-temporal variation including with music and other non-speech sounds. However, forward speech was used to compose the test stimuli so a speech mode may have been engaged by this. Also, it is possible that reversed speech is still perceived to be speech, so these studies are not conclusive of a general auditory effect.

**Tests of the effect of perceptual grouping**

It was also hypothesised that reduced boundary shifts with direction changes may be a result of perceptual ungrouping that occurs with sounds that originate from different directions. Sounds that are perceived to come from different directions may
be perceived as different auditory objects that are processed by different transmission channels. The compensation effect may have evolved to not apply compensation from one channel onto the other. Gender changes and other changes in the nature of sounds can also result in perceptual ungrouping of stimuli (Bregman 1990) and might also indicate different channels. In order to test the grouping hypothesis gender changes and changes in the type of sound (reversed/forward speech) were used between precursor and test sound. The test was run with a 160 ms time gap to engage mainly the spectral compensation effect rather than the enhancement effect. No reduction in boundary shift was found with changes in gender or between forward/reversed speech. These results indicate that the effect is not sensitive to indications that the speech has come from separate channels, except where this might be signalled by direction changes. It is also noted, but not discussed in Watkins, that ungrouping may explain the reduced boundary shift with noise precursors in this experiment (there may be ungrouping between the noise precursor and speech test sound). It is possible that although ungrouping was not shown between reversed/forward and female/male, ungrouping may have occurred where the precursor was non-speech (noise) and the test sound was speech. So a grouping explanation for the noise results is not ruled out. A test with all noise stimuli would be useful for determining whether a central component can occur with noise, as ungrouping based on programme item change would be prevented.

**Studies following Watkins (1991)**

This section describes studies by Watkins and colleagues that have examined the spectral compensation effect and its role in transmission channel compensation further.


In the previous study by Watkins the affricate ending ‘ch’ in the ‘Itch’-‘Etch’ test sounds was left unfiltered along with the test vowel. In Watkins and Makin (1996b) this was filtered with the same filter as the precursor sentence. Watkins and Makin (1996b) hypothesised that larger compensation effects would be seen when the affricate was filtered alongside the precursor stimulus because this removes ambiguity as to whether the test vowel belongs to the same transmission channel as the precursor or to the
same transmission channel as the affricate. However, because the affricate is short and involves less spectro-temporal variation this might not carry sufficient information about the transmission channel to influence the test vowel. An enhancement effect due to adaptation cannot affect preceding stimuli, so it was hypothesised that only the workings of a central spectral compensation effect would be tested here.

When filters were applied to test word endings and no precursor was present a significant boundary shift was observed. Shifts were only slightly smaller than for filtered precursors. These results show that the spectral compensation effect works backwards. It is interesting that these short endings produced significant boundary shifts. This suggests that: a) the enhancement effect is not likely to be behind these shifts, the shifts must represent a central process that can influence the perception of prior sounds and; b) sentence-length precursors are not necessary for the spectral compensation effect to calculate LTAS. However, based on Watkins’ earlier conclusion that spectro-temporal variation is necessary, sufficient spectro-temporal variation in the ending might be needed. The effect was similar for noise-like affricates ‘ch’ and periodic endings ‘pt’ but there was an increased boundary shift for the ‘Slow’-‘Flow’ continuum. The ‘Slow’ - ‘Flow’ ending contains the most spectral variation, so this might explain the increased boundary shift for this stimulus.

In this study the authors elaborated on the central mechanism proposed to be behind the spectral compensation effect; an auditory memory explanation was put forward. Spectral envelopes at different points in the signal are compared. This requires the storage of details of a spectral envelope at one point in the signal until a comparison at a subsequent point can be made: ‘The duration of such a store seems to be no longer than one or two phonemes as the size of effects on test sounds from the short subsequent sounds in (this study) are about the same as the size of effects found with longer, more complex phrases’ (Watkins and Makin 1996b). Short-term timbral memory involves detailed representations of timbre with a high enough resolution to allow listeners to distinguish between small differences in spectra (Cowan 1984). However, no further explanation of how this occurs has been given. It is necessary to further examine how the spectral compensation effect fits with what is already known about timbral memory. For the reasons mentioned in Section 2.2, it is not clear how memory has a role in causing channel compensation and the boundary shifts seen here. Future studies
should examine the role of memory in the effect further as memory is not currently known to be involved in channel compensation.


The effect of shifts in the categorisation of test words after filtered precursor sentences is similar to the effect of shifts in the categorisation of test words after sentences with raised/lowered F1 and F2 values, as in Ladefoged and Broadbent’s (1957) study. Watkins and Makin (1994) examined whether the spectral compensation effect explains Ladefoged and Broadbent’s (1957) results. Ladefoged and Broadbent (1957) declared that their effect showed adaptation to speaker ‘formant ranges’ and a speech specific mechanism (see Section 5.1.3). However, Watkins and Makin argue that the spectral compensation effect results in compensation for LTAS and shifts are the contrastive direction to this rather than in response to any individual formants or formant ranges. Compensation to LTAS may not be speech specific. The contrastive shifts in categorisation in Ladefoged and Broadbent’s (1957) study also may also be in contrast to LTAS and may be caused by the spectral compensation mechanism.

Watkins and Makin (1994) aimed to establish whether the compensation process seen in both Watkins’ (1991) and Ladefoged and Broadbent’s (1957) studies uses variation in F1 range or the LTAS of the speech. An initial experiment compared the effect of manipulating formant ranges as in Ladefoged and Broadbent’s (1957) study, and of manipulating the long-term average spectrum, via filtering, on the identity of test vowels, as in Watkins’ (1991) study. The phrase ‘Hello, you’ll hear the sound’ was used as a precursor sentence. Precursors were produced with: a low F1 range (200 Hz subtracted from each F1); a high F1 range (200 Hz added to each F1); the inverse filter of the average spectrum of /e/ or; with the inverse spectrum of /i/ applied. Test words varied along an ‘Itch’-‘Etch’ continuum. Listeners were asked to identify test words following the various precursors. The experiment confirmed the Ladefoged and Broadbent (1957) effect with F1 range-varying precursors. Words sounded more like ‘Itch’ (low F1) with high F1 range precursors and more like ‘Etch’ (high F1) with low F1 range precursors. Boundary shifts were similar to those seen when inverse filtering of the vowel spectrum was used.
In another experiment precursors were both adjusted for F1 range as in Experiment 1 and filtered with the inverse spectrum of /e/ or the inverse spectrum of /i/. The results showed that the effect of F1 range persists when precursors are played through the inverse spectrum filters but the size of effect is smaller, illustrating that much of the effect of F1 shift is accounted for by filtering for the LTAS. The authors concluded that the Ladefoged and Broadbent (1957) effect with F1 range manipulation is mediated by LTAS. Compensation is for the long-term characteristics of the signal, rather than for the formants which are time-varying characteristics.

Watkins and Makin (1994) also tested the effect of the F1 range manipulation using reversed speech precursors (this had already been done for inverse filter precursors in Watkins’ (1991) study). It was expected that if the compensation effect for the F1 range precursors operates via speech mechanisms then the effect might be reduced with reversed speech because of its ‘unnaturalness’. The shift was not reduced by reversed-speech precursors. The effect with F1 range manipulations and inverse filter precursors is similar in this respect. Neither appears to be speech specific (as was originally thought for the F1 range manipulation). This further indicates that the same process is involved in the studies by Watkins (1991) and Ladefoged and Broadbent (1957) and that it is a general auditory process. However, this test might be considered an inadequate test of speech specificity as the reversed speech may be sufficiently speech-like to engage a speech system. Further, as in Watkins’ (1991) study, the precursor was reversed but the test word was not. This non-reversed test word may have been enough to engage speech compensation mechanisms.

Watkins and Makin (1994) showed that the spectral compensation process uses the long-term spectrum of the filter and applies this in compensation. Therefore talker formant ranges, per se, are not important to speaker normalisation. The authors elaborate on the process behind spectral compensation: the current study further supports the necessity for spectro-temporal information (no central effect with speech-correlated noise). However, it revealed that although spectro-temporal variation is important the process does not use this in compensating for the transmission channel. Formants and formant ranges are not used, rather the long term average of signal is calculated and this is applied to test sounds. Further support for this is seen in studies by Holt (2006) and Laing et al. (2012) which also find that compensation is
based on the average spectrum rather than specific formants or formant variations (see Section 5.3).

**Other studies by Watkins et al.**

More recently Watkins et al. (Beeston et al. 2014, Watkins 2005a,b, Watkins and Makin 2007, 2011) have investigated the effect of temporal distortion caused by room reflections on speech identification. These studies show that prior experience with the channel (hearing prior speech processed with the same reverberation) results in compensation for the masking and temporal smearing effects caused by reflections and improves speech intelligibility. These studies are not discussed further here as they examine temporal distortion caused by reflections rather than spectral distortion. However, it is acknowledged that these studies may be relevant to understanding how colouration arriving via room reflections is compensated for, and even how colouration arriving via other channels is compensated for. It is likely that a similar mechanism is involved in removing the spectral effect and the temporal effect of reflections as both cause the same physical perturbation of the signal via an unchanging continuous (e.g. loudspeaker resonances) or repeating (e.g. room reflections) addition or subtraction of energy that is not part of the signal, but part of the channel. It is also true that this effect is similar to the effect of background noise that results in an unchanging addition of energy not belonging to the source. The research in Watkins et al’s studies on reverberation points towards *noise reduction* via the medial olivocochlear (MOC) efferent pathway, as an explanation for compensation for reverberation. Such a mechanism may contribute to the compensation of spectral colouration caused by rooms or other channels. Beeston et al. (2014) have published a model of compensation for the effects of room reverberation based on the MOC process. This model predicts the restoration of speech intelligibility in reverberation after prior listening. As well as predicting compensation for temporal smearing caused by room reverberation, this model might also predict shifts in the perceived timbre of a test sound after prior listening to the room (or other channel), if both are a result of an MOC efferent process. The time course of the MOC effect is similar to that of auditory enhancement (recovery is within 500 ms), so the MOC process may be involved in compensation via the enhancement effect.
The spectral compensation effect summary and discussion

Studies by Watkins et al. have revealed a more complicated picture of channel compensation. Watkins (1991) shows that compensation may arise from two different mechanisms. A monaural enhancement mechanism and a central mechanism which is termed the spectral compensation effect.

In Watkins’ studies the enhancement mechanisms is, again, shown to be monaural and have a short time course with a notable reduction in effect with a 160 ms time gap (a full estimate of completed recovery is not given). The tests with noise precursors appear to provide further evidence that the enhancement effect is a general auditory process occurring with non-speech sounds. However, speech was used in the test sounds so a speech specific process may have been engaged. There was some evidence that the enhancement effect is reduced in magnitude with speech precursors compared to the static noise precursors used in Summerfield’s work. The enhancement mechanism has not previously been investigated with a non-static item and a mechanism that relies on peripheral adaptation might be expected to reduce with spectral variation in the signal. If this result shows a genuine reduction in enhancement with real-speech compared to static stimuli it questions the strength of this effect to produce overshoot in real speech, where phonemes contain spectral variation. However, Watkins’ test with real-speech was conducted with a 160 ms time gap which would also have reduced the enhancement component.

When speech precursors were used a large contrastive boundary shift occurred due to a central ‘spectral compensation’ mechanism. This central effect appears to be direction sensitive and to occur only where the stimuli contain speech (though not necessarily natural speech as reversed speech is adequate), or only where they contain spectral variation. This central effect was not seen in Summerfield’s work. This may have been either because Summerfield’s stimuli were non-speech vowel complements, or because they contained no spectral variation. It is also possible that the spectral compensation effect may have come about in Watkins’ (1991) study because precursors were slightly longer, *sentence-length*, precursors compared to Summerfield’s. Watkins proposes that the effect compensates for the LTAS of the prior sentence. It may be that an effect that compensates for LTAS requires longer listening or at least a precursor of sufficient
length to achieve this properly. However, the effect was shown to occur with short sounds following the test sound so length may not be important.

There is some evidence to suggest that the amount of spectral variation in the precursor may be important and longer length may be beneficial for allowing increased sampling of this spectral-variation. Little is known about the time course of the spectral compensation effect, except that it might have a longer onset than enhancement (listeners might benefit from longer precursors), and that the recovery is shown to persist over a 160 ms time gap. This persistence beyond 160 ms silence is suggestive of a longer time course than enhancement. However, on its own this is not enough to distinguish its offset from that of the enhancement effect because in Summerfield et al.’s studies the enhancement effect was found to last up to 500 ms or longer in silence. The spectral compensation effect was shown to be the same effect as seen in Ladefoged and Broadbent’s (1957) study, which had an offset of 10 seconds, so the offset of the effect may be 10 seconds or longer. Further studies are required to measure the time course of this effect in more detail.

The spectral compensation effect appears to be somewhat reduced by direction change between the precursor and test sound. The reduction in the boundary effect with direction change supports a hypothesis that the mechanism is designed to compensate for sounds that are produced through the same transmission channel. However, the effect does not reduce with other cues of channel change, such as a change in gender of the talker or reversal of precursor speech. Watkins observed that as the effect is sensitive to direction but not other changes this might reflect the fact that a single transmission channel often involves different voices but not sources from different directions. Perceived direction change of room reflections also reduces the precedence effect and initiates a renewed adaptation-like build up of the effect (Clifton and Freyman 1994). Direction sensitivity may indicate that the two processes are related.

The mechanisms behind this effect are not certain, but a peripheral neural process is ruled out due to its crossaural nature. The effect may be cognitive and it is shown to work backwards. As a result of its central nature and longer time course, timbral memory was proposed to be involved in this effect. The effect appears to provide compensation for LTAS of a whole sentence and memory may be necessary for
calculating LTAS. However, memory is not direction sensitive so a separate cognitive mechanism would be needed to prevent compensation where direction change occurs. Further, it is not clear how memory causes shifts in perception. The spectral compensation effect appears to be the same effect behind the compensation for VT seen in Ladefoged and Broadbent (1957), where relational processing was put forward as an explanation. It is true that timbral memory is likely to be involved in relational processing, as relationships must be remembered, so the two explanations may be compatible, but timbral memory is not currently known as a mechanism of causing relative perception or channel compensation more generally. Other mechanisms may explain the spectral compensation effect. Watkins has elaborated on the importance of spectral change to the spectral compensation mechanism. Watkins indicates that a separation of channel from the source may occur because the spectral changes in speech are separated from the less rapid changes in the transmission channel. Once this separation occurs the LTAS of the precursor can be established and this is compensated for. This separation of change from the channel appears to imply a perceptual grouping mechanism behind the spectral compensation effect (a similar explanation regarding the grouping of spectral change separately from the prior spectrum was given for enhancement).

Generally the importance of spectral change fits with the observation that enhancement effect is dependent on spectral change (Summerfield et al. 1987) and also research by van Dijkhuizen et al. (1987) and Haggard et al. (1987) has shown that transmission channels with time varying responses reduced speech intelligibility, suggesting that a time varying channel could not be compensated for. However, the exact role of spectro-temporal variation in bringing about compensation mechanisms is not known. It is possible that the spectral variation in the signal is necessary for the calculation of LTAS, allowing the listener to sample the characteristics of the channel at multiple frequencies to build up a picture of the average channel spectrum (like a sine sweep moving through a channel). The requirement for spectral variation in the signal and the purpose of this needs to be further clarified. It may be that speech in the precursor rather than spectral variation is necessary for the effect.

As stated previously the enhancement effect, as seen in Watkins’ work, may contribute to compensation for both the phonetic context and the channel. Its primary role
appears to be in compensation for the prior phoneme, however, rather than the channel. The spectral compensation effect, if it calculates LTAS over a sentence length precursor, may be more suited to compensation for the channel. It is less likely that spectral compensation explains the overshoot effect as the time course may be longer than the phoneme. However, evidence of a longer time course is not confirmed, so it is possible that this effect also contributes to overshoot. Watkins’ spectral compensation effect has been shown to explain VT compensation and therefore it is likely that it also reduces colouration by loudspeakers and rooms. However, as noted above there may be differences in compensation for the VT and loudspeakers and rooms.

There are also differences in colouration that arrives from the loudspeaker and room. Watkins’ study involved the presentation of stimuli shaped by a filter and played over headphones. This closely represents the filtering that occurs to sounds via telephones and earphones and arriving directly from loudspeakers, but does not reflect the filtering that reaches the listener due to the room, whereby colouration arrives in the form of reflections from different directions. The direction sensitivity of the process may mean that compensation for the spectrum of reflections arriving from different directions does not occur (see also Watkins (1999)); the effect may not compensate for rooms as well as other channels.

The results of Watkins’ tests show enhancement with noise precursors further suggesting that is it a general auditory process. However, the use of speech in the test prevents a firm conclusion regarding speech specificity. Further, it appears that enhancement may be reduced with real-speech and is largest with static stimuli. So this effect may favour certain types of sounds. It appears that the spectral compensation effect may be speech specific as so far it has only been observed when speech precursors and test sounds are used. However, if spectro-temporal variation is all that is necessary, then other non-speech sounds may elicit the effect. The fact that it occurs with reversed speech may be evidence of a general auditory nature. However, reversed speech can still be heard as speech and in Watkins’ work all test sounds were speech and may have been sufficient to elicit a speech mode of processing.

In conclusion the results of Watkins’ work appears to show two mechanisms that may be involved in channel compensation: the enhancement effect, which may enhance a
changing sound source relative to the channel and the spectral compensation effect, which is central and may act in a similar way or provide compensation for the LTAS of a longer sample of prior speech and remove the LTAS of the channel in this way. The difference between the two effects appears to be in the time course over which they operate—one compensating for the effects of the prior phoneme and the other compensating for the effects of the LTAS (the static channel); their monaural/central origins; and the primary aims of the effect—compensation for overshoot and compensation for the channel. However, further information, primarily on time course, is necessary to confirm these roles. Also, it is necessary to show that these mechanisms occur with speech and non-speech alike if they can be regarded compensation mechanisms for channels more generally.

5.2.3 Section summary and discussion

The previous section described work to determine mechanisms of compensation for a particular type of transmission channel (the VT) and the phonetic context. The work in the current section aims to determine mechanisms of channel compensation in speech perception more generally. In this process the work additionally sheds light on mechanisms behind compensation for the VT and phonetic context. The mechanisms of channel compensation identified may be considered potential mechanisms of the time-gap sensitive component of real-world channel compensation if they are shown to be time-gap sensitive and the Step 1 Research Process Questions A-C (see Sections 5.2 or 4.1) can be answered for these mechanisms. The extent to which the mechanisms are time-gap sensitive and the extent to which Questions A-C can be answered is discussed Section 5.4. However, the mechanisms will be summarised and discussed here and answers to the chapter questions will be given.

The enhancement mechanism: Summerfield and colleagues presented listeners with flat spectrum or noise precursors filtered by a channel with the shape of a vowel complement. These precursors produced the perception of the vowels in following test sounds, thereby showing a mechanism that enhances spectral change relative to the spectrum of the immediately prior sound and compensates for the effect of filtering to the precursor. This effect is not crossaural (i.e it is monaural). The effect may be
appropriate for compensation for the phonetic context (overshoot) as it is quick acting (onset approximately 150 ms and recovery approximately 500 ms) and therefore only appears to compensate for the prior phoneme and enhances spectral change between adjacent phonemes. However, this effect is also hypothesised to remove the channel by increasing the perception of a spectrally changing source relative to the static channel, thereby increasing the signal to noise ratio of the source. Therefore, this effect may explain compensation for the VT and loudspeakers and rooms, as well as compensation for the phonetic context. The effect is small but the exact magnitude is difficult to assess across studies. It may be sufficient to explain overshoot and may be sufficient to remove overlap due to VT variation. It may provide compensation for other channels sufficient to remove colouration to a notable degree. However, it is clear that channel colouration can be heard to some extent, so any compensation via this mechanism cannot be considered complete.

The research by Watkins et al. showed two mechanisms of compensation for transmission channels: the enhancement mechanism and a newly identified ‘spectral compensation’ mechanism. An aim of Watkins’s studies was to test the role of the enhancement effect in compensation for channels generally but using sounds richer in spectrum compared to Summerfield (i.e. real-speech sounds and broadband filtering rather than notch filtering). Noise precursors filtered for the inverse spectrum of vowels showed an enhancement effect very similar to that revealed by Summerfield. The effect was monaural and not direction sensitive. The magnitude of the effect is difficult to assess and compare with Summerfield’s in a meaningful way with the boundary shift measure used by Watkins, but the effect was small. Onset time was not measured but there was notable recovery within the 160 ms gap between the precursor and test sound, which fits reasonably well to the recovery times for the enhancement effect given by Summerfield. Again, the time course demonstrated appears to show an effect that can compensate for the prior phoneme and enhance spectral change between phonemes. It appears that tests with the spectrally static noise precursors by Watkins demonstrated summerfield’s enhancement effect. Further, a monaural enhancement effect was also found with real-speech precursors (which are not spectrally static). However, this appeared to be smaller than expected showing that the enhancement effect may be reduced when the precursor is not static and may be small when listening to running
speech.

Watkins and Summerfield’s work show that the enhancement effect was eliminated with crossaural presentation. Therefore the effect appears to be monaural and a peripheral neural adaptation explanation is supported. This adaptation may be simple adaptation of the neural firing rate and/or adaptation of suppression. Watkins further confirmed that the effect is not direction sensitive, supporting this explanation. Further, both sets of studies show that onset and decay time are in line with adaptation in the peripheral hearing system. Cognitive explanations such as perceptual grouping have been put forward but these do not appear to explain some of the aspects of the effect. It has been concluded that cognitive effects may work alongside enhancement to produce constancy, but they are unlikely to cause enhancement themselves. These explanations remain possible but a peripheral neural explanation best fits the current data. However, so far it has not been possible to find consistent evidence of peripheral adaptation in physiological studies. Adaptation processes occurring higher in the hearing system are beginning to be investigated as explanations of the enhancement effect, including adaptation arising in the MOC pathway.

The spectral compensation mechanism: when Watkins tested enhancement using real-speech sentence-length precursors, rather than the static noise and harmonic complex precursors used by Summerfield, a significantly larger boundary shift indicative of compensation occurred. Like with noise precursors the shift reduced with crossaural presentation suggesting the enhancement effect was partly behind this. However, the effect was not entirely reduced suggesting a central mechanism also contributes to compensation in this case. Further, the component of the boundary shift that was non crossaural was tested to determine whether a central but direction sensitive (rather than monaural) process was behind the effect. There was a reduction in boundary shift when peripheral prior stimulation was equivalent at both ears (as in the monaural studies), but there was a perceived direction change between precursor and test. Therefore, part of the non crossaural effect appears to be caused by a central direction sensitive process, rather than monaural process. A small portion of the reduction in the crossaural tests was not explained by this, and appears to be monaural.

The origins of the boundary shift with real-speech precursors is therefore more complex
than with noise precursors. There may be: a) a central process that provides a boundary shift even when presentation is crossaural or at both ears but from different directions; b) a central but direction-sensitive process that provides a shift when stimulation is equivalent at both ears, but reduces as the direction change gets larger and; c) a monaural process that provides a portion of the boundary shift only when stimuli are presented at the same ear. a) and b) may be a single central direction-sensitive process, but where even a full direction change of 180 degrees does not fully break the central effect and some boundary shift will occur. c) indicates the presence of a separate mechanism that is monaural and appears to represent the contribution of the enhancement effect. However, Watkins claims that the monaural effect in this study is too small and short-lived to be important to channel compensation. The results indicate that enhancement is smaller when sounds other than the static stimuli are used. This may be expected by a peripheral adaptation mechanism, as adaptation may occur most where stimuli are static. However, this test contained a 160 ms time gap, which appears to have contributed to the reduced enhancement effect. It would be interesting to conduct the same speech test with this time gap removed to determine the real relative contribution of the central and monaural mechanisms with real-speech stimuli, however this is not conducted as part of this thesis.

The exact time course of the spectral compensation effect was not established in these studies but in Watkins’ (1991) study it was shown to persist over 160 ms, unlike the enhancement effect in that study. This persistence does not definitively distinguish the time course of spectral compensation from the enhancement effect, because in Summerfield et al’s studies the enhancement effect was found to last 500 ms or longer in silence. It appears that the effect onsets with the sentence length precursors used, but sentence length precursors may not be necessary for the effect to occur as Watkins and Makin (1994) showed the effect with phoneme-length sounds following the test sound. However, it has been shown that the effect appears to cause boundary shifts in respect of the LTAS of the whole prior sentence, so it appears that a full sentence length precursor is required for compensation in respect of this LTAS. Overall, it can be concluded that the evidence suggests that this is a compensation mechanism with a longer time course than enhancement and thereby it applies compensation for the
LTAS of a longer segment of prior speech.

The underlying mechanisms behind the spectral compensation effect were not investigated but, as the effect is central and backwards acting, memory has been suggested as a cause. Memory would not explain the direction sensitive nature of the effect but a separate mechanism working alongside memory, may prevent the same compensation being applied to sounds from different directions. The involvement of memory was suggested by the authors to account for the fact that a picture of LTAS over a longer time period must be stored for shifts to occur in respect of this. However, it is not clear how memory would cause the shifts in the spectra of test sounds described. As noted in Chapter 1, memory appears to be a process by which noise in judgements is increased/decreased, rather than a process that causes perceptual shifts in a particular direction as seen here. A longer onset neural adaptation process that occurs centrally may also be a cause but this was not discussed. A monaural peripheral neural adaptation explanation is ruled out as this cannot act backwards and is not crossaural.

Speech specificity: The enhancement effect appears to occur in masking studies, studies with vowel complements, static flat spectra, and noise precursors. This suggests that speech is not necessary to elicit the effect. However, all studies specifically testing enhancement, rather than masking, have used speech sounds somewhere in the test, which may bring about a speech mode of listening. Only FSV studies have not used speech sounds but a perception of a vowel is elicited in these studies. Only masking studies provide a truly non-speech test, but methodological differences mean it is possible that masking studies show an effect that is different to Summerfield’s enhancement effect. It is possible that the effect at work with the sine tone precursors and narrow noise bands used in masking studies is specific to simpler sounds, such as two-tone suppression. It is therefore not absolutely certain that enhancement as shown in Summerfield’s work and Watkins work with spectrally rich sounds occurs in an entirely non-speech scenario. Enhancement is likely to be a general auditory process, however, as there is strong evidence that it occurs via peripheral neural adaptation, which is obligatory, acting on any sound passing through the ear. However, this process may still favour some sounds (e.g. static sounds, like noise) over others (e.g. spectrally varied sounds, like music and speech). Further, the peripheral nature of the process is
not firmly confirmed, so it remains possible that the effect is speech specific or at least different with speech.

The spectral compensation effect appears to be engaged where real-speech but not noise is used as a precursor to ‘carry’ the channel colouration. The spectral compensation effect may have been engaged with real-speech precursors because it is a speech specific mechanism. The effect occurred when the precursor was reversed suggesting that the listener does not need to hear the precursor as natural speech, but in this case the listener can still recognise the reversed speech as speech and a speech mode of perception may occur. Watkins also hypothesised that this effect occurs whenever there is spectro-temporal variation in the signal; the real-speech precursor contained this but the noise did not. Given this, the spectral compensation effect may not be speech specific. However, further tests are necessary to confirm this. Another possible explanation is that the slightly longer precursors used in Watkins’s work compared to Summerfield’s brought about the spectral compensation mechanism in this study. However, length appears to not be important in bringing about the effect. The length of the precursor was not much longer than those used in Summerfield’s work and Watkins and Makin (1994) show that short phoneme-length consonants following the test vowel bring about the spectral compensation effect. The effect with these phonemes was larger where there was more spectro-temporal variation supporting a hypothesis that length or speech may not be important to the spectral compensation effect but sufficient spectro-temporal variation may be.

**The role of the mechanisms in channel compensation:** The role of the enhancement effect in compensation for general transmission channels, the VT and the phonetic context was discussed. The effect is shown to enhance spectral change relative to static energy. It appears that the effect may enhance spectral change leaving the channel relatively reduced in perception. The quick acting nature of this effect appears to make it suited for enhancing changes between phonemes and explaining overshoot. As explained, overshoot may cause channel perception in the same way. The processes of enhancement and overshoot appear to be related and may be the same mechanism; both may cause channel compensation in the same way. However, Watkins (1991) declared the enhancement effect to be too small and short-lived to be responsible for significant transmission channel compensation. Watkins’ main arguments against
5.2 LITERATURE REVIEW PART 2: GENERAL COMPENSATION FOR THE CHANNEL IN SPEECH PERCEPTION

Enhancement’s role in channel compensation are that is it quick to recover and may recover with gaps in running speech and it is not direction sensitive, so inappropriate compensation may be applied to sounds arriving from different channels. Also the effect was particularly small in tests with real-speech precursors. However, this result may have been partly due to the 160ms gap, which reduced the effect. According to Watkins the effect appears to be more appropriate to explaining the perceptual overshoot effect, but he does not rule out a role for this effect in channel compensation. The enhancement effect does not appear to be an effect specifically designed to remove the channel, but to create overshoot. But as noted for overshoot, it may contribute to compensation for channel as well as the prior phoneme.

The general nature of the spectral compensation effect is similar to the enhancement effect: contrast between adjacent sounds is enhanced. This nature suggests that the effect may have a role in the perceptual overshoot effect. However, it appears that it may have a longer time course and provide compensation for the LTAS of longer (sentence-length) precursors rather than the immediately adjacent phoneme, when this longer context is available; thus speaker variation, but not phoneme effects, are compensated for. However, the time course of the spectral compensation effect has not been confirmed. The experiment by Watkins and Makin (1994) confirms that the spectral compensation effect is the same effect at work as in Ladefoged and Broadbent’s (1957) study and that both studies appear to provide compensation for the LTAS of the whole prior sentence. An effect that compensates for the LTAS of a longer segment of prior speech (or other programme item) appears ideally suited to compensation for the spectral effects of a static channel.

The direction sensitivity of the spectral compensation effect may mean that it is suited to producing compensation for the channel as this means that the same compensation is not applied when a channel change is cued by a direction change. However, a very quick acting effect like the enhancement effect will also prevent the old compensation from being applied to a new channel. The spectral compensation effect may require spectral variation. The importance of spectral variation is not clear but it appears that it may allow for a calculation of the LTAS to be applied in compensation or this variation may somehow aid separation of the spectrally changing sounds from the static spectrum (i.e. an ungrouping between the channel and the source). It is apparent
that a sampling of a number of prior speech sounds made by the same talker, each with different spectra, may help the listener obtain a picture of the channel response. This might be akin to passing a sine-tone sweep through the VT and advantageous to obtaining a picture of the impulse response of that system. However, the requirement and role of spectral variation in instigating the spectral compensation effect is not clear. Tests to determine whether spectral variation, rather than speech, is a requirement for the spectral compensation effect should be run using Watkins’ (1991) paradigm and non-speech sounds. If similar effects occur with non-speech (and continue to be absent with static sounds like noise) then it can be concluded that spectro-temporal variation is likely to be necessary rather than real speech. Such tests are conducted as part of this thesis and reported in Chapter 6.

Chapter questions

The work in this section has contributed to answering the chapter questions described at the beginning of this section.

**Question 5.1** asked: a) To what extent does compensation for the spectral effects of the vocal tract occur in speech perception? b) What are the mechanisms behind this? c) What is the time course of this compensation? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

The enhancement effect may be a mechanism of VT compensation. This appears to enhance spectral change in a source relative to the static channel, thereby removing the effect of the channel relative to the spectrally changing source within the channel. This effect is similar to the overshoot effect and both may arise from the same underlying mechanism. Both Summerfield’s work and that of Watkins et al. show that the enhancement effect is monaural and a peripheral adaptation explanation is likely based on this and its short time course. However, a centrally produced, monaural (e.g. efferent) process is not ruled out and no firm evidence of its peripheral nature has been provided. Cognitive mechanisms such as auditory grouping and attention have not been ruled out. Watkins concludes that this effect may only be a small contributor to channel
compensation when listening to real speech because its short time course means that the compensation would cease with relatively short silent gaps in a running programme item and the magnitude of the effect is too small to offer notable compensation for the channel. The mechanism may be considered time-gap sensitive as it recovers with approximately 500 ms of silence.

The spectral compensation effect appears to be more suited to compensation for the VT. Watkins and Makin’s (1994) study showed directly that the spectral compensation effect was behind the VT compensation seen in Ladefoged and Broadbent’s (1957) study. It appears that in both cases the LTAS of the prior sentence is compensated for, thus removing the spectrum of the channel from the source. Other features of the spectral compensation effect, such as its direction sensitivity and time course, appear to be appropriate for channel compensation. A sentence-length precursor may be necessary. This would ensure that the compensation is for channel LTAS rather than the phoneme, but this is not confirmed. Phoneme length precursors also appear to bring about the effect (Watkins and Makin 1994). It may be that only sufficient spectral variation within the precursor is necessary for this effect to onset. The recovery time appears to be longer than enhancement. If the mechanism is the same as that in Ladefoged and Broadbent’s (1957) study it may break after 10 seconds. It is therefore time-gap sensitive and will not apply compensation for a precursor to a sound where there is a silent gap between them. However, recovery time needs to be confirmed.

Enhancement studies and spectral compensation studies together show that mechanisms behind VT compensation may be both monaural (possibly peripheral) and central, and to some extent direction sensitive (possibly central neural adaptation or memory based). Both are time-gap sensitive. The enhancement effect appears not to be speech specific as there is some evidence of this occurring with sine-tones and noise in masking studies. However, studies specifically investigating this effect as a means of channel compensation have used speech stimuli somewhere in the test so the general auditory nature of enhancement is not confirmed. Speech anywhere in the test may elicit a speech mode of listening. The spectral compensation effect has so far only been tested where speech is present either in the precursor or test sound, so a general auditory process is not confirmed. However, if it is found that this mechanism simply requires spectral variation in the precursor, as is suggested by Watkins, then this may
be a general auditory process (see the answer to Question 5.4 for more on the speech specificity of these effects).

**Question 5.2** asked: a) To what extent is the phonetic context compensated for? b) What are the mechanisms behind this compensation? c) What is the time course? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

Overshoot was described in Section 5.1.3 as a mechanism that primarily compensates for the phonetic context. This section reveals that the enhancement effect is similar to the overshoot effect and may be the cause of overshoot, although this was not specifically suggested by the authors of the enhancement work who describe this effect simply as a mechanism of channel compensation. Further work would be desirable to directly test whether enhancement is a mechanism of compensation for the phonetic context, and whether it explains overshoot. Further evidence contributing towards the conclusion that enhancement and overshoot are the same process is presented in the next section of the literature review (see Section 5.3). The enhancement effect appears to be particularly well suited to be a mechanism for compensation of the phonetic context as its time course is short. The role of the spectral compensation in compensation for phonetic context is not clear. The time course appears to be longer than enhancement and it appears that compensation is for the LTAS of a longer sentence. In which case, in running speech, this will remove the effects of the channel but not the prior phonetic context. This appears to make this mechanism unsuitable for explaining overshoot. The time course, time-gap sensitivity and general auditory nature of these effects were discussed in the answer to Question 5.1.

**Question 5.3**: asked a) To what extent is the effect of transmission channels other than the vocal tract (e.g. loudspeakers, rooms) compensated for in speech perception? b) What are the mechanisms behind this? c) What is the time course of effects? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

The work presented does not involve the examination of compensation for the phonetic context or VTs specifically, but compensation for filtering more generally when listening
to speech. The filtering is more representative of filtering caused by generic channels such as loudspeakers and rooms because it is wideband. However, the filtering was not especially representative of that caused by some channels, such as loudspeakers, because the filtering was for inverse vowel spectra. In Summerfield’s work this consisted of notch filtering to represent formant peaks. Watkins’ studies use smoother filter transfer functions compared to Summerfield’s but the filter is still inverse-vowel shaped and so still contains more spectral peaks (where formants occur) than would be produced by loudspeakers and probably also rooms. This filtering may therefore be considered more extreme than that caused by loudspeakers and rooms. The effects of loudspeakers and rooms and similar channels may be milder but it is expected that the enhancement effect and spectral compensation may both play a role in compensation for this milder filtering (though the requirement for these mechanisms may be less strong). The time course of enhancement is quick but that of spectral compensation may be longer. As mentioned above, a longer time course may be more suitable for compensation for static channels as the LTAS will be used in compensation. The direction sensitivity of the spectral compensation effect may mean that it is less suitable for compensation for the room. Mechanisms behind these effects, time course, time-gap sensitivity were discussed in the answer to Questions 5.1 and 5.2.

**Question 5.4** asked: a) To what extent does compensation for transmission channels occur when listening non-speech sounds? b) What are mechanisms behind this? c) What is the time course of effects? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time gap sensitive?

All work described here used speech sounds somewhere in the test and this may have brought about a speech mode of listening, this limits conclusions regarding the general auditory nature of effects. The enhancement effect was shown here to occur with noise precursors but the use of speech in the test sounds mean that a general auditory process is not confirmed. However, in some of the research presented in this section an enhancement-like effect is shown to occur in non-speech tests with noise and tones (i.e. masking tests) but these may elicit different effects to the channel compensation effects seen here. There is also strong evidence that it occurs via low level neural adaptation, so the balance of the evidence currently suggests that it is not speech specific. If caused by peripheral adaptation it may be larger with static stimuli compared to stimuli with
spectral variation, such as speech. There is evidence to suggest that it is smaller with real-speech compared to noise. The fact that the spectral compensation effect does not occur with noise suggests speech specificity but the spectral compensation effect is shown not to be unique to natural-sounding speech sounds, occurring when speech precursors were reversed. It may simply require spectral variation. If the spectral compensation effect only requires spectral variation it may occur with music and may be a general auditory process.

5.2.4 Section conclusion

The work in this section contributes to answering Research Question 1b, and Step 1 Questions A-C. The enhancement effect and the spectral compensation effect appear to show mechanisms that allow a listener to compensate for spectral colouration caused by transmission channels, including loudspeakers and rooms. The enhancement effect works by increasing change relative to the immediately prior spectral context. It works over a short time course with phoneme-length onset and recovery periods. There is some evidence to suggest that this effect and the overshoot effect are the results of the same mechanism. Therefore the overshoot effect may also be regarded as a channel compensation process. The spectral compensation effect appears to also enhance the perception of change but relative to the LTAS of a longer (sentence-length) sample of the spectral context. It appears to be similar to the extrinsic VT compensation mechanism discussed in Section 5.1 so appears to also have a longer recovery period than enhancement. The spectral compensation mechanism which provides compensation for the LTAS of the longer prior auditory context seems particularly suited to channel compensation. These mechanisms are shown to be time-gap sensitive and may have the required nature and time-course to offer real-world compensation. The extent to which this is the case (i.e. the extent to which Questions A-C for can be answered for these mechanisms) and whether they can be considered potential mechanisms of the time-gap sensitive component of real-world compensation for loudspeakers and rooms will be discussed further in Section 5.4. Questions A-C must be answered for these mechanisms when listening to non-speech as well as speech if they are to be considered general auditory compensation mechanisms that may affect
listening to any sound through loudspeakers and rooms. The extent to which these mechanisms occur when listening to non-speech sound remains a question of primary importance to this thesis and this is investigated in the next section of the literature review and via experiments presented in Chapter 6.

5.3 Literature review part 3: General compensation for the channel in non-speech perception

The studies presented in the previous two sections have aimed to investigate channel compensation in speech perception. They have not specifically investigated the extent to which the compensation mechanisms are general auditory processes. Studies have mainly used speech stimuli or have elicited perceptions of speech in the test. It is clear that some speech perception mechanisms are likely to be general auditory or general cognitive processes (e.g. peripheral neural adaptation, gestalt perception, categorical perception), but there is much research suggesting that speech perception benefits from a link with the physical aspects of articulation (motor theory hypothesis), or that a *speech-mode* of perception exists, whereby different mechanisms are engaged when listening to speech (Fowler 2006, Liberman et al. 1967, Repp 1982). Therefore, some mechanisms may be specific to speech perception. Timbral constancy and mechanisms that compensate for the channel may be different when listening to speech and non-speech.

This section primarily aims to determine the extent to which the previously discussed mechanisms cause channel compensation when listening to non-speech sounds and whether any additional compensation mechanisms are seen when listening to non-speech. However, it will be seen that most of the studies that purport to investigate whether the compensation mechanisms are general auditory mechanisms also involve speech being present in the test and therefore may not be good tests of the general auditory nature of the compensation mechanisms. However, these studies shed light on non-speech perception and are discussed here. Further, these studies also contribute
to determining the mechanisms behind the compensation effects discussed previously. A few studies have investigated perceptions in tests that only use non-speech sounds (e.g. music) and these will also be described in this section.

Therefore, like the previous sections of the literature review, this section contributes to establishing a list of mechanisms that might be potential mechanisms of the time-gap sensitive component of real-world compensation for loudspeakers and rooms (like the previous sections, this section will contribute to fulfilling Step 1a(i) of the research process—see Chapter 4). Those mechanisms discussed here which are time-gap sensitive and for which Step 1 Questions A-C can be answered (see below and Chapter 4), will be considered potential mechanisms of the time-gap sensitive component of real-world compensation for loudspeakers and rooms. If a number of such mechanisms can be found then Research Question 1b will be answered. The previous parts of the literature review have discussed such mechanisms in speech perception (contributing towards fulfilling Step 1a(i) of the research process—see Chapter 4, for speech listening). However, it was noted in Chapter 4 that compensation mechanisms in both speech and non-speech listening should be found so that potential mechanisms of compensation for loudspeakers and rooms when listening to any sound are established. Therefore the specific aim of the current chapter is to determine a number of such mechanisms that occur in non-speech perception (contributing fulfilling Step 1a(i) for non-speech listening).

Step 1 Questions:

A) For any compensation mechanism, does the mechanism produce a measurable effect that is indicative of compensation and compatible with providing the real-world compensation seen in Olive et al.’s (1995) experiment and in the experiments described in Chapter 3?

B) Is there one or more distinguishing feature of this mechanism (e.g. is it crossaural or does it require a long onset)? This is required to:

- confirm that the mechanism is a unique and distinct mechanism and not a manifestation of another compensation mechanism and
• separate this compensation mechanism from other compensation mechanisms in tests of elimination in Step 2.

C) Is the time course of the mechanism such that it is compatible with being an explanation of the real-world compensation seen in Olive et al.’s (1995) experiment and in the experiments described in Chapter 3?

To guide the collection of the information necessary, the chapter questions will be addressed. These are shown below. The answers to the chapter questions will be given in the summary and discussion of this section, see Section 5.3.5). Particular emphasis will be placed on answering Question 5.4:

**Question 5.1:** a) To what extent does compensation for the spectral effects of the vocal tract occur in speech perception? b) What are the mechanisms behind this? c) What is the time course of this compensation? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

**Question 5.2:** a) To what extent is the phonetic context compensated for? b) What are the mechanisms behind this compensation? c) What is the time course? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

**Question 5.3:** a) To what extent is the effect of transmission channels other than the vocal tract (e.g. loudspeakers, rooms) compensated for in speech perception? b) What are the mechanisms behind this? c) What is the time course of effects? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

**Question 5.4:** a) To what extent does compensation for transmission channels occur when listening non-speech sounds? b) what are mechanisms are behind this? c) What is the time course of effects? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

A few researchers have aimed to examine whether channel compensation mechanisms are general auditory processes. Holt notes a division in this research between two types
of studies. There is research concerning ‘auditory processing that is sensitive only to interactions among temporally adjacent acoustic events’ (i.e. research that examines the effect of local spectral context) and research concerning how listeners are ‘responsive to distributional regularities among multiple acoustic events that unfold across time’ (the global spectral context) (Holt 2006). Primarily, research with non-speech stimuli has investigated the effect of local context on the perception of test sounds. This research aims to determine whether overshoot is a speech specific process. Because overshoot may be involved in transmission channel compensation, these studies may show that this source of channel compensation is not speech specific. Researchers have also examined the effect of the global context on perception of non-speech test sounds. These studies have investigated channel compensation that occurs via mechanisms that compensate for the LTAS of the longer prior context (i.e. a spectral compensation-like process, as in Ladefoged and Broadbent (1957) and Watkins’ (1991) studies). The discussion of work in this section will therefore be divided into studies that examine the overshoot effect (compensation for local spectral context) and studies that examine compensation for the LTAS of longer precursors (or global context).

As mentioned above the tests described have used a mix of speech and non-speech sounds. Only a couple of studies have tested compensation in an entirely non-speech context. The review of studies is therefore split as follows:

1. Studies that examine the effect of local context using mixed speech and non-speech sounds (Subsection 5.3.1).

2. Studies that examine the effect of global context using mixed speech and non-speech precursors and test sounds (Subsection 5.3.2).

3. Studies that use entirely non-speech stimuli (Subsection 5.3.3).

A summary and discussion subsection will discuss the extent to which the chapter questions have been answered (see Section 5.3.4). The extent to which Research Question 1b can be answered using the information presented in this section is discussed in the conclusion to this section (see Section 5.3.5), in brief, and in more detail in Section 5.4 of this chapter and in Chapter 7. The answers to Step 1 Questions A-C, and which of these mechanisms can be confirmed to have the potential contribute to the
time-gap sensitive component of real-world compensation for loudspeakers and rooms will be discussed in Section 5.4.

5.3.1 Mixed speech and non-speech tests—local context effects

A number of experiments have investigated the effect of phoneme-length non-speech precursors on the categorisation of speech test sounds in order to determine whether overshoot is speech specific and whether enhancement or spectral compensation is involved in overshoot. These studies therefore examine local context effects on the perception of test sounds (Coady et al. 2003, Holt and Lotto 2002, Holt et al. 2000, Holt and Stephens 2003, Lotto et al. 2003, Lotto and Kluender 1998), or see Lotto and Holt (2006) and Holt and Kluender (2000) for a review.

Non-speech analogues of ‘VC’ syllables within ‘VC-CV’ stimuli, have been tested for their effect on the perception of following syllables. These non-speech precursor sounds represent the ‘minimal characteristics of their speech equivalents’ (Holt and Lotto 2002). Lotto and Kluender (1998) tested a non-speech precursor analogue of the perceptual overshoot study by Mann (1980) (see Section 5.1.3). Precursors consisted of sine-wave transitions to model the F3 transition and offset frequency of /al/ and /ar/ syllables. Test sounds were real-speech syllables that varied perceptually between /ga/ and /da/. The same shifts indicative of overshoot were seen in this study as in the original all-speech study (Mann 1980). The boundary shifts were of a similar magnitude in both experiments.

Lotto and Kluender (1998) also showed a spectrally contrastive categorisation shift when the precursors were single non-linguistic sine-tone transitions modelling F3. Another example can be seen in Holt’s (1999) study, where non-speech equivalents of vowel precursors /i/ or /u/ (i.e. a single sine tone transition that mimics the F2 transition in /i/ and /u/ and distinguishes between these vowels) caused shifts in speech consonants varying perceptually from /ba/ to /da/. In this study the non-speech glides shared the spectral qualities of formant cues in real precursor vowels but were reported to be neither acoustically nor perceptually like real speech, suggesting that non-speech
context has an influence on vowel identification. Again, the size of the boundary shift was similar with speech and non-speech precursors.

The role of enhancement and spectral compensation in overshoot

These non-speech precursor tests suggest a general auditory process is involved in overshoot. They may also show that enhancement or spectral compensation is behind overshoot. Holt & Lotto (2002) observed the similarity between overshoot studies and auditory enhancement as seen in Summerfield’s work: ‘It has been suggested that spectral contrast in speech may be a special case of auditory enhancement or at least that the mechanisms responsible for the two processes are the same’.

There is already good evidence that enhancement is not speech specific and that enhancement occurs with non-speech precursors (Summerfield et al. 1984, Viemeister 1980, Watkins 1991). So if enhancement is behind overshoot it is not surprising that overshoot occurs with non-speech precursors. The non-speech precursor tests may be further evidence of enhancement working to cause overshoot. It is less clear whether these studies show a central process working to produce overshoot. These non-speech precursor tests were not conducted crossaurally. The studies described in the previous two sections show that speech precursors may be necessary to elicit the spectral compensation effect. However, spectral variation in the precursor, rather than speech may be sufficient and the current tests did contain spectrally varied precursors. It is possible that the central spectral compensation effect as seen in Watkins’ (1991) work, or another central effect, is elicited.

To determine the centrality of the effect Holt and Lotto (2002) conducted an all-speech crossaural test. This test showed overshoot with contralateral presentation of the precursor and test sound, demonstrating that a central effect must be involved in overshoot. However, boundary shifts were smaller than with monaural presentation, indicating some role for a monaural process alongside the central effect. Holt and Lotto (2002) concludes that both central and peripheral processes are involved in overshoot: ‘[spectral compensation] and enhancement may work together in producing effects of contrast in audition [...] enhancement and [spectral compensation] may originate from distinct mechanisms yet both serve the common purpose of promoting contrast’ (Holt
This result appears to match that of Watkins (1991) where shifts in response to sentence-length precursors also consisted of a monaural and a central component. Therefore, Holt and Lotto’s (2002) work may be seen as extending that of Watkins (1991) to show the spectral compensation effect with phoneme-length speech precursors. This test may be evidence that both the enhancement effect and Watkins’ (1991) spectral compensation effect are involved in overshoot and that the spectral compensation effect does not require a sentence length precursor. This is not entirely surprising as Watkins and Makin (1994) have already showed the spectral compensation effect with phoneme length sounds. However, Holt and Lotto (2002) also investigated the recovery of the central component of overshoot, showing that this breaks with only 275 ms of silence. Watkins’ work suggests that recovery of the spectral compensation effect is longer than enhancement (as measured in that study) and Ladefoged and Broadbent (1957) shows that it might last up to 10 s in silence. Generally speaking a longer time course would be expected for a mechanism that compensates for the LTAS of prior sounds. It is therefore possible that the short time course central effect seen in Holt and Lotto’s (2002) study indicates a separate effect, which is also central but has a quicker time course than the spectral compensation effect. Such an effect could contribute to overshoot alongside enhancement as its time course is appropriate for this. Alternatively, the central effect seen here may be the spectral compensation effect and the short time course may be caused by the fact that the spectral compensation effect is weaker with shorter precursors and therefore offset is quicker. If Holt’s work does show the spectral compensation effect then this is evidence that there is no central component to overshoot. This is because, although the spectral compensation effect would be evident where precursors are short and shifts would be in response to the LTAS of the prior phoneme (giving a central version of the enhancement effect), where a longer context is available, the spectral compensation mechanism would calculate the LTAS of this longer context and apply this in compensation instead. This would mean that in running speech compensation for the channel, but not the phoneme, is provided. Therefore, the central overshoot effect seen here may be an artefact of the study design and occur due to the presentation of a precursor phoneme in isolation, not within running speech. Further, research would be needed to confirm that it is the
spectral compensation effect causing the central effect in this overshoot study, not a separate short-time course central effect.

**Speech specificity of overshoot mechanism**

Lotto et al. (2003) conducted the same test with non-speech precursors mimicking F2 transitions in al/ar and measured the effects on ga/da perception. The same pattern of results was found: both a monaural and a central effect contributed to overshoot. Effects were similar in magnitude as with the all-speech study. The only difference was that the recovery time for the central effect was slightly shorter in this test than in Holt and Lotto’s (2002) study, occurring with 175 ms of silence. This may show that the effect with non-speech precursors is slightly less robust than in the all-speech test by Holt and Lotto (2002), but Lotto and Holt conclude that the central effect is equivalent with speech and non-speech and suggestive of a central general auditory compensation mechanism behind overshoot:

> “the agreement of the current results with those of Holt and Lotto (2002) implicates similar mechanisms in both kinds of [speech and non-speech] context effects. This agreement can be added to the mounting evidence for a general auditory role in speech context effects”

This work may show that the central effect in overshoot (i.e. the spectral compensation effect) occurs to the same extent with non-speech stimuli. However, Lotto et al.’s (2003) test may not be an adequate test of non-speech perception as the non-speech formant transitions may have elicited speech perceptions. Sine-wave transitions are often perceived as speech (Remez et al. 1981) and may bring about a speech mode even where the listener does not perceive the sounds to be speech-like (Stilp et al. 2010). Therefore, the similarity between overshoot with speech and non-speech might be due to the fact that both tests engaged a speech-mode of perception. In Watkins’ and Summerfield’s work where the phoneme precursors were more clearly not like speech (noise and flat spectra filtered for vowel complements) no crossaural effect was seen. Even if the precursors do not elicit a speech mode of perception, the real-speech test sounds might be sufficient to do this. Therefore, the crossaural contribution in Lotto and Holt’s study may be a speech effect.
There are some issues arising from previous studies that count against the argument that the central effect is speech specific, however: a) while only studies using speech precursors have shown a central effect, all of these studies have also contained spectral variation and it could be this that causes the effect rather than speech, b) Watkins’ tests with noise show that speech in the test sound is not sufficient to induce the central effect, so the mere fact that a speech test sound is used is unlikely to bring about a speech mode, but where precursors are speech-like and the test sound is real speech, this combination is likely to be adequate to bring about a speech mode if one exists. In conclusion, a speech mode cannot be ruled out when either the precursor or the test sound is speech.

There is also evidence from overshoot studies against a role of enhancement in overshoot, or at least enhancement caused by simple adaptation in the auditory nerve. A couple of studies have not found a neural basis for perceptual overshoot using recordings at the auditory-nerve and cochlear-nucleus (peripheral sites) (Holt and Rhode 2000). It would be useful to conduct future research to confirm those findings and also to obtain physiological evidence for other proposed peripheral cases of overshoot and enhancement, such as adaptation of suppression and the MOC efferent mechanism.

It can be concluded that the mechanisms behind overshoot are still not confirmed. The monaural component of the effect appears to be of the same nature and time course as enhancement (though the peripheral origins of overshoot have been questioned by Holt and Rhode (2000)). It can be said with some certainty that the enhancement effect contributes to overshoot, but the mechanisms behind enhancement need further research. A central effect may be engaged alongside a monaural effect. However, this central effect is likely to be the spectral compensation effect working with short precursors. If so, even though a central effect is seen in overshoot tests there may be no central component to overshoot in running speech, if the central effect compensates for LTAS of a longer spectral context. If, however, the central effect seen in these studies is a separate effect with a short time course, then this newly discovered effect may contribute to overshoot. If overshoot contributes to channel compensation via enhancing change relative to the static channel, then this new central effect may be a source of channel compensation.
Regarding speech specificity, if the precursors generate non-speech processing then these studies may show that enhancement and spectral compensation (or a new central effect) occur with non-speech. It is not surprising that this is the case for enhancement. But these studies would constitute the first studies showing spectral compensation with non-speech precursors. They may therefore support a conclusion that spectro-temporal variation in the precursor is sufficient to bring about the spectral compensation effect. However, speech in the test sound and the possible interpretation of the sine tone precursors as speech limit this conclusion, as does the fact that the central effect may not be the spectral compensation mechanism but a separate central effect.

It may be concluded that the enhancement effect is likely to be a component of overshoot (if not the sole mechanism behind overshoot). This may provide compensation for the channel via overshoot. Lotto et al.’s (2003) work may provide further evidence that this occurs when listening to non-speech, but the presence of speech test stimuli in those tests and the possibility of the sine tone precursor being heard as speech, means that this conclusion is not strong. A central effect may cause overshoot, but this is unlikely because the central effect seen in Holt and Lotto (2002) and Lotto et al.’s (2003) work is more likely to be a manifestation of the spectral compensation effect, which will not cause overshoot. Either way, the central component is not conclusively shown to occur with non-speech because of the presence of speech test stimuli in those tests and the possibility of the sine tone precursor being heard as speech, so Lotto et al.’s (2003) study is unlikely to show that the spectral compensation effect is a general auditory process.

**Speech precursors and non-speech test sound**

A couple of studies have examined compensation with non-speech test sounds with speech phoneme-length precursors. If shifts occur, this is further evidence towards a general auditory process behind overshoot, but not conclusive because speech is present in the precursor. From the studies described so far it appears that the nature of the precursor is important in bringing about compensation effects and the test sound is just a recipient of the compensation effect. However, it has also been argued that speech in either the precursor or the test sound may engage a speech mode of processing.
A test by Holt and Stephens (2003) used speech precursors and non-speech test sounds and found boundary shifts in the perception of the test sounds. However, it is not clear whether a central mechanism or enhancement was behind this effect as no crossaural test was conducted. A larger effect was found when a test was run with all speech sounds. The authors claim that this was because the speech test sounds were easier to identify than the non-speech sounds, rather than any added compensation effect being brought about by the speech test sound. However, it is possible that this result was because a stronger (possibly central) speech specific mode of compensation was elicited in the all-speech case. This would suggest that without the speech test sounds only the smaller enhancement effect was elicited. However, ungrouping between a non-speech precursor and speech test sound may also explain reduced shifts in that case as was proposed for Watkins’ enhancement tests with noise precursors and speech test sounds. Tests with only non-speech stimuli that use the same type of stimuli are preferable to examining whether compensation occurs via a general auditory process.

Mixed precursors, local context summary

Studies that have investigated the effect of non-speech, phoneme-length precursors on overshoot have aimed to determine the mechanisms behind overshoot and the extent to which these mechanisms occur with non-speech sounds. Further evidence for the role of a monaural mechanism in the overshoot effect was provided. Overshoot occurs, at least in part, via a monaural process and the time course of overshoot matches that of enhancement. This shows that enhancement is a likely cause of overshoot. Enhancement is already not thought to be speech specific, but the current studies may show further evidence of this, if the precursors used can be considered ‘non-speech’.

The studies also show that a central effect may be involved in overshoot. This may be a specific central effect with a short time course that can play a role in overshoot. However, the central effect may be a manifestation of the spectral compensation effect, with short precursors. The spectral compensation is thought to build up a picture of the LTAS of the longer context, but where only a short context is available it may compensate for the LTAS of this and act like the enhancement effect. If so, the central effect seen in overshoot studies is not likely to be involved in overshoot for running speech, as the spectral compensation effect will not compensate for the
phoneme in running speech. However, the assumption that the spectral compensation effect compensates for LTAS of longer precursors still needs to be confirmed.

If the studies can be regarded as non-speech tests then these studies show that both the monaural and the central process is non-speech specific. These studies might therefore show that enhancement and spectral compensation are not speech specific. However, the presence of speech in the test sound limits this conclusion as this may be sufficient to bring about a speech mode, and further, the sine-tone transition precursors may have been interpreted by the hearing system as speech and may also elicit a speech mode of perception. A test without speech is necessary to confirm that overshoot is a general auditory process. Studies with speech precursors but non-speech test sounds provide little evidence of speech specificity but provide further evidence that ungrouping between precursor and test sounds may prevent compensation effects. Future experiments should use the same programme item for precursor and test sound to avoid ungrouping.

5.3.2 Mixed speech and non-speech tests—global context effects

The studies discussed in the previous subsection examine the effects of local context via phoneme-length precursors but not the global context. Some authors acknowledged that sensorineural systems are responsive to change and track reliability in the environment across longer time periods (Coady et al. 2003, Holt et al. 2001, Kiefte and Kluender 2008). Extending the length of the precursor may increase the strength of compensation or a separate compensation mechanism may be evident after a threshold listening period.

Based on the experiments described above, evidence that longer listening benefits compensation is not particularly strong. The enhancement effect has been shown to be maximal with phoneme length precursors, so a larger effect with longer precursors is not expected (but it is possible that a different monaural mechanism may begin after some threshold listening period). Longer listening has also been shown to be unnecessary for the central spectral compensation effect to occur. Holt and Lotto (2002), Lotto
et al. (2003) and Watkins and Makin (1996b) appear to show this with only phoneme length precursors. However, the nature of this effect may change with longer listening, e.g. providing channel compensation rather than phoneme compensation, or it may become larger or more robust to silent gaps with longer precursors.

Where longer than phoneme precursors have been tested previously, precursors were still quite short (1.2-1.4 second long sentences were used in Watkins’ studies and approximately 2 second long sentences were used in Ladefoged and Broadbent’s (1957) study). Differences in effects with longer, rather than shorter, precursors are not clear. However, based on the research presented so far it may be concluded that there appears to be a mechanism called the spectral compensation effect, which compensates for the LTAS of the whole sentence-length precursor rather than the immediately prior spectrum (Watkins and Makin 1994). This requires a longer time course than enhancement and is central (Ladefoged and Broadbent 1957, Watkins 1991, Watkins and Makin 1994). Further confirmation of such a mechanism is provided by work by Holt as is the non-speech specificity of this mechanism. The work by Holt is described in the following paragraphs.

Holt (2005) put forward an acoustic histories hypothesis: ‘context need neither be linguistic or local to have an effect on speech categorisation’. Hearing a randomly ordered sequence of sine tones, by sampling a range of the spectrum, will influence categorisation of a following speech sound and provide channel compensation.

Holt (2005) used a non-speech representation of sentence-length precursors which consisted of a train of 21, 70ms sine tones of different frequencies presented across 2.2s with 30ms gaps between tones. This acoustic history precursor was followed by a final neutral standard tone (70ms, 2,300Hz) and a silent period of 50ms. A test sound continuum was made using male spoken natural speech whereby onset frequencies of F2 and F3 were varied to create a series of stimuli that spanned the endpoints of /ga/ to /da/. The mean frequency of acoustic histories corresponded to the F3 offset frequencies of /al/ and /ar/ precursors, which have been shown to produce shifts in the categorisation boundary for a /ga/ to /da/ test sound continuum (Mann 1980). The mean frequency of the distributions was either 1,800Hz (range 1,300-2,300Hz with tones varying in 50 Hz steps) or 2,800Hz (range 2,300-3,300Hz, 50 Hz steps)
corresponding to /ar/ and /al/ F3 offset frequencies respectively. Sequential acoustic characteristics of the tone histories were varied on each trial but LTAS was the same. The standard tone was expected to have no effect on labelling the /ga/ to /da/ targets but prevent any boundary shifts originating from the immediately adjacent context from affecting the test stimulus.

More /ga/ syllables were identified after the presentation of a history with the higher frequency mean. More /da/ syllables were perceived after a history with the lower mean. The effect was largest in the middle of the /ga/-/da/ range, where test sounds were perceptually ambiguous. The effect withstood the neutral standard tone, showing that the effect extends beyond temporally adjacent stimuli. A condition where the mean of the histories was held constant but the range was varied did not cause boundary shifts. This study provides evidence that the LTAS of a non-speech precursor can shift the perception of speech categories in a contrastive manner as seen in previous studies. Shifts appear to be in contrast to the mean spectrum of the history, suggesting that the whole prior context before the immediately adjacent sound is used in the shift, not simply the immediately prior context. Further, a neutral standard tone and silent gap does not prevent the mean history producing the boundary shifts but these do prevent adjacent context affecting the test sounds. However, it is possible that the effect may have been reduced by this neutral standard and silent gap somewhat if perception is weighted toward the local spectrum. Holt also tested the effect of extending the silent period after the standard tone from 50 to 1,300 ms. Contrary to what might be expected if a peripheral adaptation process was behind the results, there was a larger boundary shift with an increase in the silent interval. This test provides clearer evidence that the whole prior spectrum is used in compensation, not just the temporally local section, but the window length is still unknown. Holt observed that the:

‘persistence of the acoustic histories effect is significantly longer than has been observed for previously-reported effects of temporally adjacent precursors [...] from what is presently known, the effect of acoustic histories on speech categorization appears, from its time course, to be distinct from the effect of temporally adjacent contexts’ (Holt 2005).

The acoustic histories effect was not tested crossaurally but the nature and time course
of the effect (compensation in light of the LTAS of a long precursor rather than adjacent context, and compensation in spite of a standard tone/silent gap) supports a conclusion of a central rather than peripheral process. Holt claims that the acoustic histories effect cannot be due to peripheral adaptation, as stimuli must be adjacent to be affected by peripheral neural adaptation. In spite of this claim, while a peripheral contribution is ruled out, it is not certain that a monaural process is not behind this effect. Crossaural tests should be conducted to confirm that this effect is central. However, current evidence that the effect is central is strong.

Holt (2006) conducted another study to determine the extent to which distribution characteristics, other than the mean of the acoustic history, play a role in the effect. Holt (2006) examined the effect of changing the mean, the variance, and the density of the acoustic history sequence (more tones closer to the mean frequency) and the number of tones in the history. Additionally, the relative role of local and global spectrum on boundary shifts, was examined and the effect of spectral complements (notched noise) rather than sine tones was tested.

As expected, results with notched noise complements of tone precursors caused shifts in the opposite direction to the tone precursors, confirming that the average spectrum of the acoustic history drives the compensation effect. There was no effect of changing the variance and effects are not affected by the density of the sampling distribution of the tones. Holt noted that this is surprising due to density effects in vision where ‘changing the variance of the colours in a test spot’s surrounding, while holding the mean constant, induces a contrastive shift in the perceived contrast and saturation of the test spot’s color’. (Holt 2006). However, changes in variance would not be expected to be relevant if it is LTAS that drives the mechanism. Histories with varying numbers of tones were tested (3, 5, 6, 9, 11, 21, or 26 tones; history duration 300-2,600 ms). The length of the sequence co-varied with the number of tones.

The magnitude of the effect of the acoustic history mean on speech categorisation increased with the number of tones composing the distribution. The results appear to be similar to those seen with adaptation to reflections in the build-up of the precedence effect (Clifton and Freyman 1994) whereby increased reflections bring about increased compensation for the effects of reflections on localisation. It is interesting
that an increase in the number of tones in the history has been reported to increase compensation. An increase in history should result in an increase in certainty as to LTAS. This result may imply that as the picture of LTAS becomes more certain the compensation becomes stronger. However, it would be expected that the magnitude of compensation would be more closely related to the size of the distortion, than the certainty about the nature of distortion. This finding may simply show an initial build-up effect.

Further evidence showing the longer time course of the effect was provided. Histories were composed to consist of sub-segments with different mean spectra. Conditions with a final high mean segment (LMH/MLH) were compared against those possessing a final low mean segment (HML/MHL). Final segment mean did not have an effect on phoneme categorisation. Compensation was clearly for the LTAS of the whole precursor. This finding could mean that in running speech the effect of local precursor spectrum does not occur. The effect of global spectrum overrides this, and perceptual overshoot effects do not occur (i.e. shifts do not occur in response to local precursor spectrum, only in response to LTAS). However, in this study (and in Holt’s (2005) study) a lack of local effect does not show a lack of the effect of the immediately prior spectrum—there was no immediately prior spectrum here because this was replaced by a tone and silence. Instead, this study shows a lack of effect of the more recent spectrum. Therefore, a role for the immediately adjacent spectrum is not ruled out. When this has not been controlled for, the overshoot effect may work alongside compensation for the LTAS seen here. Holt notes:

‘this result is not incompatible with Lotto and Kluender’s (1998) finding that adjacent non-speech tones shift /ga/-/da/ categorization as a function of tone frequency because the acoustic histories of the present experiment were separated from the speech syllable by a constant mid-frequency standard tone.’

Therefore, further evidence is provided that when a longer context is available LTAS is compensated for, but a separate mechanism that compensates for the immediately prior context is not ruled out. Holt further speculates that for the effect of local spectrum, peripheral adaptation might still occur alongside the acoustic histories effect:
‘peripheral adaptation might be responsible for the effects of single temporally adjacent speech and non-speech context stimuli on speech categorization’.

It would appear that overshoot must occur in running speech (where LTAS is also available), as overshoot appears necessary to prevent formant overlap caused by co-articulation. It is unlikely that when a longer context is present overshoot does not occur. However, the operation of this alongside compensation for LTAS has not been examined. Further work should examine how these effects interact. However, Holt’s (2006) results do appear to show that there is not a gradual weighting towards the most recent spectrum.

The time course and apparent central nature of the acoustic histories effect in both Holt (2005) and Holt’s (2006) studies shows that the effect is similar to the spectral compensation effect. This similarity could be because the effects originate from the same mechanism. The effect seen in Ladefoged and Broadbent’s study, was also shown to be caused by the spectral compensation effect (Watkins and Makin 1994) and therefore the effect in Ladefoged and Broadbent’s study, the acoustic histories effect and the spectral compensation effect may all be manifestations of the same mechanism. In all cases effects appear to be in response to the LTAS of the whole sentence length precursor, rather than the immediately adjacent portion of this: they are produced centrally and have a time course longer than enhancement. A study by Huang and Holt (2012) was conducted to confirm that the acoustic histories effect is the same effect seen in Ladefoged and Broadbent’s (1957) study.

As in Watkins and Makin’s (1994) study, Huang and Holt (2012) sought to determine whether the acoustic histories effect explains the results of Ladefoged and Broadbent (1957) by determining whether the same effects would occur with sine tone precursors as with the real speech precursors with modulated F1 and F2 used in the original study. An acoustic histories sine tone sequence to represent the long-term talker characteristics and /bet/ and /but/ test words identical to the original study were used. Boundary shifts similar to those in Ladefoged and Broadbent’s (1957) study were found with sine tone precursors. Again, context was shown to have a greater influence on perceptually ambiguous targets. Speech precursors like those in Ladefoged and Broadbent’s (1957) study were also tested. There was no interaction between context
type (speech versus non-speech) confirming that there was no difference in boundary shift when tone histories or real speech precursors were used. Huang and Holt (2012) conclude:

‘a representation of the ‘talker’ is not necessary—instead of solving the inverse problem to recover the actual neutral vocal tract shape as an articulatory referent for normalization, listeners may use the average spectrum LTAS as an auditory referent [...] Talker normalization is likely to be a multi-facet phenomenon [...] Nonetheless LTAS, which provides sufficient context information via general perceptual processes, is an important factor in the adaptive coding of speech perception processing that contributes to talker normalization.’

Therefore, the findings provide strong evidence that Ladefoged and Broadbent’s effect, the acoustic histories effect, and the spectral compensation effect are caused by a single mechanism. It is concluded that Holt’s acoustic histories effect and Ladefoged and Broadbent’s (1957) effect are manifestations of the spectral compensation effect. Watkins’ term is used as this was the first effect specifically proposed to compensate for the LTAS of a longer sample of prior speech.

Holt’s results appear to show the spectral compensation effect occurs with non-speech sounds. The acoustic history was not heard as speech, nor can it be said to share features with speech other than spectro-temporal variation. However, there is speech in the test sound so a speech mode of perception cannot be ruled out. The compensation process would need to be instigated before the listener knows they are listening to speech (i.e. before they hear the test sound), so a speech mode explanation is less likely. However, the speech in the test sounds leaves a possibility for an effect that only occurs once the listener knows they are listening to speech. Therefore, future work remains necessary to confirm that the spectral compensation effect is not a speech specific process. Tests without speech should also be run to confirm the spectral compensation effect is a general auditory process.

Assuming a single central spectral compensation effect is shown in all global context studies, the exact mechanisms behind this effect need to be determined. Because of
the longer time course and apparent compensation for LTAS it has been predicted that memory is involved in boundary shifts with longer precursors (Watkins and Makin 1994, 1996b). Holt (2005) states that: ‘this context effect requires some degree of memory as the acoustic history unfolds in time and appears to indicate a degree of abstraction from the precise acoustic signal presented on any given trial’. As well as memory, Holt (2005) also puts forward a central adaptation hypothesis, specifically Stimulus Specific Adaptation (SSA) within the central auditory cortex (Ulanovsky et al. 2004, 2003). It is proposed that the role of the primary auditory cortex may be to respond to regularity in the long-term statistics of auditory signals by depressing neural response to regularity, and enhancing response to auditory novelty. This is achieved by SSA. SSA occurs in populations of fibres at many locations within the central auditory system from the mid-brain to the primary auditory cortex. The extended time course of the central effect may be evidence of SSA. Holt (2006) notes that: ‘at the periphery, neurons show somewhat weak, fast adaptation but adaptation becomes progressively stronger, with longer time constants along the ascending auditory pathway.’ The longer time course seen with the spectral compensation effect is therefore compatible with adaptation effects at higher sites: ‘At the algorithmic level, SSA is perhaps better suited than adaptation-by-fatigue to relate to context-dependent auditory processing of the sort investigated here’ (Holt 2006). However, cognitive mechanisms such as memory can also involve longer time courses.

There is good evidence that the mechanism of boundary shifts with longer precursors is the same in all cases, and that it may be said that Watkins’ (1991) spectral compensation mechanism is behind all centrally produced boundary shifts discussed in this literature review. To confirm this, ideally, the similarity of effects across studies should be further investigated. It should be confirmed that all effects thought to be a manifestation of the spectral compensation effect involve a central mechanism (as shown by Watkins (1991) and indicated by Holt (2005, 2006), and not tested for in Ladefoged and Broadbent’s (1957) study), can be distinguished from enhancement based on their use of the LTAS, rather than the immediately adjacent spectrum (as seen in Holt’s (2006) study and indicated by Watkins (1999) and Ladefoged and Broadbent (1957)), and that all effects occur with non-speech precursors (as suggested in the studies by Holt (2005, 2006), hypothesised but not tested by Watkins and not tested
Another test of global context with non-speech sounds was conducted by Kluender and Stilp (2006). In this test speech precursor sentences were filtered for the inverse spectrum of either a French horn or a tenor saxophone and test sounds were from a Horn-Sax continuum. A compensation effect was seen but, again, the speech precursor may have brought about a speech specific mechanism. This compensation also appears to be in response to LTAS rather than the immediately adjacent spectrum, which suggests a central effect is involved, but this is not tested with crossaural tests.

It is possible that an increased tendency to hear horn/sax after an inverse horn/sax precursor could result from the effect of filtering on only the immediately prior spectrum. Therefore, it is possible that this effect is only demonstrating enhancement. The experiments were also conducted with precursors and test sounds originating from different directions. The effect was not diminished with direction change. This further indicates that the spectral compensation effect was not the cause of these results and possibly only the enhancement mechanism caused shifts in this test. This test may therefore provide evidence of enhancement occurring with non-speech (if it can be considered a non-speech test) but does not appear to suggest that the central compensation mechanism occurs with non-speech.

Global precursor effects summary and discussion

Boundary shifts indicative of compensation were examined with precursors longer than phoneme length, which were composed of non-speech sounds (Holt 2005, 2006). These studies are non-speech equivalents of sentence-length precursor studies by Watkins (1991), Watkins and Makin (1994), Watkins and Makin (1996b) and Ladefoged and Broadbent (1957). As in those studies, a shift in test sounds the opposite direction of filtering to a precursor spectrum occurred. Shifts in response to LTAS of a whole sentence-length precursor rather than the local spectrum was confirmed by the fact that: the nature of the shift shows a shift that is in response to the LTAS of the whole precursor and cannot be in response to any other aspect; the LTAS of the final segment of the precursor before the test sound does not affect boundary shifts; the shifts still occur when the immediately prior sound is held constant and there
is a time gap of over 1.3 seconds and the range of the spectrum is not important to the effect. The fact that there was no effect of the nearest precursor segment suggests that there is no weighting towards the more recent spectrum in perception but there may still be an effect of immediately adjacent context when this is not controlled for. Studies on co-articulation suggest that in running speech, compensation for the immediately adjacent spectrum is necessary, so these shifts are expected to occur alongside compensation for LTAS. Further tests where the immediate context is not controlled for should be conducted to confirm that both effects occur in running speech. It is not clear how the shifts in response to immediately adjacent context and in response to LTAS will interact in perception but it appears that separate mechanisms are necessary to produce both. Two potential independent mechanisms appear to exist: a central mechanism appears to compensate for the LTAS of the global context and this may provide compensation for a static channel, and a monaural (possibly peripheral) mechanism appears to compensate for the local context and provide overshoot.

The current evidence that the acoustic histories effect and the spectral compensation effect arise from the same central mechanism, is strong: both have been shown to cause the effects seen in Ladefoged and Broadbent’s (1957) work which measures compensation for the VT (Huang and Holt 2012, Watkins and Makin 1994). To support a conclusion that the spectral compensation effect is behind all centrally produced shifts seen in the studies discussed, it should be confirmed that in all cases: compensation is brought about by a central mechanism; compensation is for the LTAS and; compensation is or is not speech specific. Based on Holt’s work it is likely that the effect is not speech specific but tests with entirely non-speech test sounds should be used to confirm this. The precursors contained spectral variation and this may be sufficient to bring about the effect. Additionally, the fact that a non-speech precursor caused shifts in a speech test sound in this study suggests that ungrouping due to differences in programme item does not explain the lack of compensation seen in previous tests with non-speech precursors and speech test sounds (or vice versa).

The mechanisms that have been put forward to describe the central spectral compensation effect are neural and cognitive. SSA is a process that causes adaptation of neural firing rate in response to stimuli regularly occurring over time, this occurs in populations of neurons at multiple points in the central hearing system (Ulanovsky
et al. 2004). Memory has been posited as another explanation as it appears that a representation of the channel is built up over time and memory may be necessary to achieve this. So far timbral memory is not considered to be a form of adaptation and adaptation appears to explain the relative perception seen in these effects, so its role in compensation is in doubt. However, a number of authors have suggested that this may play a role in channel compensation so this mechanism warrants further attention.

5.3.3 Non-speech tests

An investigation of compensation in a test without speech is desirable for ascertaining compensation mechanisms with non-speech. Only a few studies described so far have examined compensation in tests using only non-speech sounds (including masking studies and Summerfield et al’s FSV tests, but FSV tests still elicited speech perceptions). If shifts occur in an entirely non-speech test this confirms a general auditory process of compensation.

The evidence presented so far suggests that enhancement is likely to occur in a non-speech context, but this is still not confirmed. Evidence largely suggests that the spectral compensation effect, however, is speech specific. Stilp et al. (2010) note that with the noise precursors in Watkins’ (1991) study, only small boundary shifts representative of an enhancement-like effect were seen but that Holt (2005) found boundary shifts in response to non-speech sine-tone precursors indicative of a spectral compensation-like effect. According to Stilp et al. (2010) the process behind compensation with non-speech test stimuli is far from clear. The authors also acknowledged that more research with non-speech is necessary to strengthen the claim that compensation is a general auditory process, and to strengthen an auditory colour constancy analogy.

The study by Stilp et al. (2010) adopted Watkins’ paradigm and used only non-speech sounds. A musical precursor phrase (1.00s of a string quintet) and a speech precursor sentence (a male voice saying ‘You will hear’), were filtered with the inverse spectrum of either a French horn or a saxophone. Test stimuli were sounds along a French horn to saxophone continuum. Precursors filtered to emphasise spectral characteristics
of a French horn elicited more saxophone identifications and precursors filtered to emphasise a saxophone resulted in more French horn identifications. This result shows that compensation occurs in an entirely non-speech test. Again, it appears that the shift is for the LTAS. However, the operation of the central spectral compensation effect is not confirmed because a crossaural test was not conducted. Precursors were longer than a phoneme but only 1 second in length so not clearly sentence-length and there was no silent gap reported between precursor and test so it is likely that only shifts in response to the immediately adjacent spectrum were sufficient to cause horn/saxophone perceptions and enhancement was largely or entirely behind the shifts. However, the precursors did contain spectral variation suggesting that the spectral compensation effect may be elicited. Shifts were similar in magnitude for both music and speech precursors. The spectral compensation effect was shown to be larger than enhancement in Watkins’ work. This may suggest that only enhancement brings about the shift and the similarity in effect size across speech and non-speech precursor conditions is because ungrouping between the speech precursor and musical test sound brought about a smaller effect in that case. Although Watkins (1991) showed that potential ungrouping between female/male speech and reversed/forward speech precursor and test sounds does not affect the spectral compensation effect, ungrouping may be more likely with speech/music than with the items that are all heard to be speech. However, Holt’s work appears to show that such ungrouping does not prevent the spectral compensation effect working with a non-speech precursor and speech test sounds. It is also possible that both conditions brought about by the spectral compensation effect. The central nature of the effect rather than the size is the decisive factor in determining whether effects are brought about by enhancement or spectral compensation. Therefore, to confirm that a central effect occurs in a non-speech test, a test with crossaural presentation should be conducted. Ungrouping should also be prevented.

An experiment by Sjerps et al. (2011) involved testing manipulated speech signals, which were modified so that they were less like speech acoustically. These stimuli were spectrally rotated speech sounds which were rated to sound unlike speech by listeners. Sentence-length precursors caused a small central compensation effect, which occurred with a 500 ms gap between the precursor and test sounds. However, this effect was much reduced compared to that seen with real speech and could have been due to
continuation of peripheral effects beyond the time gap, as compensation with crossaural presentation was not tested. Sjerps also tested more extremely manipulated speech stimuli (spectrally rotated plus pitch flattened, syllable reversed, and time-gap-removed speech). These stimuli were also confirmed to sound unlike speech. No compensation was elicited with these stimuli, except a very small effect for stimuli that underwent all the above mentioned manipulation expect spectral rotation, which again could be due to peripheral effects. Ratings of how speech-like the stimuli sounded were not clearly related to objective speech-likeness (i.e. the degree of manipulation of the speech signal). For example stimuli with rotation plus all other manipulations were rated as slightly more speech-like than the precursors with only spectral rotation. However, there was some evidence of a relationship between the extent of manipulation and compensation. The most manipulated stimuli showed no compensation, precursors that had either all other manipulations but no spectral rotation or spectral rotation but no other manipulation both showed small compensation and, as seen in Watkins’ studies, speech showed large compensation. The authors concluded that there appears to be some evidence for a relationship between stimuli being acoustically (objectively) more like natural speech and compensation, but no relationship between perceived speech-likeness and compensation. The authors observed that naturalness of the signal, rather than speech-likeness, might be important for compensation and specifically the signal should have not merely spectral variation but natural patterns of spectro-temporal variation in order to invoke learnt compensation mechanisms. Peripheral compensation may also have been less likely to occur with these distorted stimuli. It appears that compensation would be more likely to occur with familiar and natural non-speech sounds than was seen in this study. Only unnatural and unfamiliar stimuli were examined in this study, so this was not tested. Supporting this conclusion, the authors observed that ‘Any sound structure that shows LTAS constancy over time (and for which it would be beneficial to normalize, e.g., due to overlapping sound categories) could, in principle, evoke learned normalization processes.’

Therefore, any absence of compensation with the manipulated speech ‘non-speech’ items in this study may have been due to the unnaturalness of the stimuli rather than the fact that the stimuli were not perceived as speech. The evidence for small compensation with the most natural of these non-speech items is suggestive
that this is the case. However, that small compensation may have been due to a peripheral compensation process. It would be useful to examine compensation with more natural and familiar non-speech stimuli from the real-world. Music may elicit large compensation. Crossaural tests would confirm that any such compensation is central rather than peripheral. Tests of real-world musical non-speech stimuli with crossaural presentation are described in Chapter 6 of this thesis.

Further evidence of boundary shifts with non-speech can be found in Lotto et al.’s (1997) study. Japanese quail (Coturnix Japonica) were trained to peck when hearing /ga/ or /da/ followed by speech /al/ and /ar/ precursors. These creatures cannot rely on a speech processing mode or experience with speech sounds (Lotto et al. 1997). Though, again, its not clear whether the enhancement or central mechanism is working in this non-speech situation, as this test provides no evidence of a central effect.

There are also studies which have not looked at overshoot or transmission channel compensation but have examined the levels of processing in speech perception. These studies have used selective adaptation to determine the level at which different speech processes occur. If one process adapts another then this suggests that the two processes operate at the same stage of processing. These studies aim to determine the extent to which concepts (for example musical timbre) are abstract or not. This line of research could provide information that is relevant to the work of those studying transmission channel compensation but to date there has been no overlap between the fields. There are a number of differences in the aims and methodology of the work, which prevent the direct application of findings across fields. However, the work is superficially similar and may be informative of potential channel compensation processes. Two pieces of research relevant to the work described in this literature review are described below. These studies have shown that boundary shifts can occur due to adaptation (the fatiguing of a response) at an abstract level as well as at an acoustical level with non-speech sounds and this adaptation of abstract auditory objects may also be involved in channel compensation.

Samuel and Newport (1979) found that voiced speech adapts whispered speech and vice versa. The effect of adaptation is to produce contrastive shifts in perception. Whispered speech is acoustically unlike spoken speech so an abstract rather than
acoustic representation must be being adapted. Samuel and Newport (1979) also found adaptation with non-speech adaptors and speech test sounds that did not overlap in frequency content but belonged to the same class of sounds. They claim that perceptual commonalities between the stimuli, but not acoustic overlap, must be causing this effect. These results suggest that the effect of non-speech on speech sounds and vice versa might come about through interactions at a higher level than auditory processing. This may call into question the assumption that shifts in stimuli are caused by acoustic attributes or even mode engaged when hearing speech. The shifts may be a result of an even higher level process. The results suggest that shifts may be to do with more abstract cognitive processes, a speech mode that processes phonemes at an abstract level (e.g. as categories) may be involved in such adaptation.

Pitt’s (1995) work has approached the spectral context effect issue from an information processing perspective. The research seeks to determine at what level of processing sounds interact. Pitt tests this in relation to auditory processing. He describes the aims of his study as ‘undertaking another test of the proposal that abstract acoustic information is encoded at a central level in the information processing system’.

An 11-step trumpet-piano test continuum was created. Perceptually the continuum sounded like one instrument timbre transitioning into the other, but endpoints were heard to be clearly either a piano sound or trumpet sound. Precursors were trumpet or piano stimuli from the end-points of the continuum. Either the piano or the trumpet end-point 311 ms sound was played 60 times with a 650 ms gap between presentations. The listener then identified test sounds varying from trumpet to piano as either one of the two. Contrastive shifts in perception were observed following adaptors. Contralateral presentation of the endpoint adaptors and test sounds also produced adaptation that was reduced by half. This may show the central spectral compensation effect occurring in a test without speech, as the reduction is similar to that seen in previous contralateral tests. However, the paradigm is too different to that used by Watkins (1991), Summerfield et al. (1984) or Holt (2006) to confirm that the same process is occurring. In Pitt’s study different precursors were used, which gave the perception of auditory objects (pianos/trumpets) rather than the inverse of these. These give the listener abstract representations of familiar auditory objects and the
authors suggest that these abstract representations acted as adaptors and cause shifts in the opposite direction of these objects in other test sound objects. The adaptation was not in response to the acoustical attributes of the sound. The work shows that a central portion of the shift appears to be fully explained by this abstract adaptation hypothesis, as the crossaural effect still occurred where precursors were acoustically different to the test sound but had the same auditory object identity. It is surprising that acoustic overlap is unnecessary and this suggests an entirely non-auditory abstract process is behind the shifts. It is not clear if and how any acoustical overlap that appears to have occurred in the tests also contributed to shifts, but the results appear to show a different mechanism of boundary shift compared to those discussed previously. This mechanism is at least different to enhancement but it remains possible that the spectral compensation effect works at an abstract level, particularly if it is speech specific. So these results may be indicative of this effect. Stimulus and paradigm differences are too large to draw conclusions regarding the crossaural nature of the spectral compensation effect with non-speech from this work. However, this study may show that the spectral compensation effect is either a high level process and occurs with non-speech or a separate mechanism of boundary shifts that acts at a high level in the stimulus-to-perception chain, also may cause boundary shifts. This result indicates that compensation mechanisms may work at multiple different levels within the stimulus-to-perception chain and this work may represent compensation at yet another level of processing. Research in the spectral compensation field and this field should be brought together in future. However, confirmation of any compensation that might occur at the acoustic level, peripheral and central, should be confirmed with speech and non-speech before an investigation of compensation at an abstract level is undertaken.

5.3.4 Section summary and discussion

The previous section described work to determine mechanisms of compensation for general transmission channels in speech perception. Two mechanisms that may play a particularly important role in compensating for the spectral effects of the channel were described: the enhancement effect and the spectral compensation effect. These
mechanisms appear to explain overshoot and VT compensation, respectively and both appear to contribute to channel compensation more generally. However, in the previous sections these effects had only been tested in tests involving speech. The extent to which these mechanisms occur with non-speech was not clear.

The current section primarily aimed to determine the extent to which the mechanisms discussed previously cause channel compensation when listening to non-speech sounds and whether any additional compensation mechanisms are revealed when listening in an entirely non-speech context. The work described in this section contributes to determining whether Questions A-C from Step 1 of the research process (see the introduction to this section and Chapter 4) can be answered for any of the compensation mechanisms when listening to non-speech. The answers to these questions will be considered in Section 5.4. The work presented in this section also aims to provide answers to the chapter questions and in particular Question 5.4, which asks about the extent to which the mechanisms described in the literature review occur with non-speech sounds. The answers to these questions are given below. However, it was seen that most of the studies that purported to investigate whether the compensation mechanisms are general auditory mechanisms contained speech within the test or involved other limitations which mean that results are not conclusive regarding the general auditory nature of the mechanisms. Therefore, limited information has been provided to answer Chapter Question 5.4 and Step 1 Questions A-C for non-speech. However, the studies did shed some light on this issue and on the time course of, and the underlying mechanisms behind, the compensation effects discussed previously within this literature review. Further, evidence was presented that suggests a possible additional compensation mechanism: a short time course central mechanism. These findings are summarised discussed before the chapter questions are answered.

**Compensation for local non-speech context:** in previous sections it was noted that overshoot may be involved in compensation for the channel and that enhancement may be behind this process. In the current section studies which investigate overshoot with crossaural listening were described. It was confirmed that a portion of the overshoot effect is monaural with a short time course, supporting a conclusion that the enhancement mechanism causes overshoot. Therefore, overshoot and enhancement can be regarded as the same process. As well as providing compensation for the
phonetic context, overshoot/enhancement may provide compensation for the channel by reducing the perceptual importance of the channel relative to that of spectral change. In the previous section the possibility of enhancement being reduced and nearly absent with speech compared to static precursors was discussed. However, a significant monaural effect occurred with the non-static real-speech phonemes in overshoot studies, confirming that this effect can occur in running speech. No new information was discovered regarding the specific time course of the monaural (enhancement) component of overshoot/enhancement: it was confirmed to occur with phoneme length precursors but recovery of this effect was not investigated. The underlying mechanisms behind overshoot/enhancement were not further investigated, work is still necessary to confirm that this is a peripheral adaptation process. This work is outside the scope of this thesis. For the purposes of the work here it is concluded that the enhancement/overshoot process is likely to be a peripheral neural adaptation process (e.g. simple adaptation or adaptation of suppression), but that neural adaptation that occurs at a higher level (e.g. as part of the MOC pathway) may also be behind this.

The work in this section also showed that a portion of overshoot appears to be central. This central effect was not tested in prior overshoot tests and was not previously seen in enhancement tests. This component may be brought about by the spectral variation now present in the speech precursor that was not present in enhancement tests which used spectrally static sounds. This central component of overshoot may show another effect that compensates for the immediately prior spectrum and may be involved in channel compensation via enhancing spectral change. However, this central effect is probably a manifestation of the spectral compensation effect as seen in Watkins’ (1991) study, working with phoneme rather than sentence-length precursors. It is possible that where there is only a phoneme-length precursor and silence before this, then the spectral compensation mechanism produces compensation for the LTAS of the phoneme and enhancement-like shifts but in running speech, where a longer context is available, the spectral compensation effect provides compensation in respect of the LTAS of the longer context instead. Therefore, the central effect seen in overshoot studies may not actually contribute to overshoot in running speech. It is not surprising that phoneme-length precursors bring about the spectral compensation effect as this has been shown to occur
with short sounds in a test by Watkins and Makin (1994). However, it is also possible that the central effect seen in these studies is not the spectral compensation effect but a new central effect with a short time course which is appropriate to produce overshoot and channel compensation via enhanced spectral change in speech sounds relative to the channel. The central component of overshoot offsets with 275 ms of silence for speech precursors and 175 ms for non-speech sine tone transition precursors. This is a quicker recovery than expected for the spectral compensation effect based on recovery reported in Ladefoged and Broadbent's (1957) study. However, recovery of the spectral compensation effect may be expected to be quicker where only a phoneme elicits the effect, rather than a sentence length precursor, as the effect is shown to be weaker with shorter precursors (Holt 2006). Further work would be necessary to determine whether alongside a short time course monaural effect (enhancement/overshoot) and a longer time course central effect (the spectral compensation effect), a short-time course central effect contributes to channel compensation. Such work is outside the scope of this thesis. For the purpose of this thesis is it concluded that the short time course central effect seen in overshoot studies is a manifestation of the spectral compensation effect with phoneme-length precursors.

It is claimed by the authors (Holt and Lotto (2002) and Lotto et al. (2003)) that these studies show that the monaural and central components of overshoot are not speech specific as they occur with sine-tone transition precursors. However, the precursors may have brought about a speech mode of listening due to their similarity to sine-tone speech and speech test sounds were used, so overshoot (and therefore enhancement and spectral compensation) are not confirmed to be a general auditory process through this work. No other research was found that specifically investigates the effect of the immediately prior (phoneme-length) spectral context on the perception of test sounds using non-speech.

**Compensation for global non-speech context:** the effect of global speech spectral context on the perception of test sounds is demonstrated in the studies by Watkins (1991) and Ladefoged and Broadbent (1957), which show compensation for the LTAS of sentence-length precursors. The effect of global non-speech spectral context on speech test sounds is demonstrated by the *acoustic histories* effect. The acoustic histories effect shows compensation for the LTAS of a *sentence-length* train of non-speech sine
5.3 LITERATURE REVIEW PART 3: GENERAL COMPENSATION FOR THE CHANNEL IN NON-SPEECH PERCEPTION

Holt’s studies (Holt 2005, 2006) provide further evidence that a mechanism that causes shifts in response to the LTAS of sentence-length precursors, rather than the more local spectrum, occurs. This compensation for LTAS of a longer segment is confirmed by the fact that: shifts occur even when the adjacent phoneme cannot affect the test sound (i.e. where there is a standard tone and a 1.3 s silent gap); the final segment mean (before the standard and silent gap) does not influence the shift and the tone range is not relevant to shifts. The effect is very likely to be central as it has an onset and offset time considerably longer than that expected for a peripheral mechanism. A crossaural test would be beneficial to confirm this but is not necessary.

It appears that Holt’s acoustic histories effect is a manifestation of the previously reported spectral compensation effect with non-speech precursors. Support for this comes from Huang and Holt (2012), which confirms that the acoustic histories effect and that of Ladefoged and Broadbent (1957) are the same. Given that Watkins and Makin (1994) confirmed that the spectral compensation effect and Ladefoged and Broadbent’s (1957) effect are the same, all three effects appear to be caused by the same mechanism. This mechanism may be called the ‘spectral compensation mechanism’. As the spectral compensation effect has now been shown twice to be the same effect as in Ladefoged and Broadbent’s (1957) study (Huang and Holt 2012, Watkins and Makin 1994), this mechanism can be concluded to show compensation for vocal tract colouration. It is also likely to show compensation for other transmission channels. It would be beneficial to make further direct comparisons of the nature, magnitude and time course of Ladefoged and Broadbent’s (1957) effect, Watkins’ effect and the acoustic histories effect in a single study to confirm that compensation effects across these studies are the same, but for the purpose of this thesis it is concluded that these studies all show the spectral compensation effect.

Given that Holt’s acoustic history studies are an example of the spectral compensation effect, the time course of the spectral compensation mechanism has been further described. It is confirmed to have an onset period longer than the phoneme. Onset is confirmed to occur over the course of the whole sentence-length precursors used (2.2 s). Offset is greater than 1.3 s of silence plus a 70 ms standard tone. Complete onset and offset was not measured so the time course may be even longer than suggested by Holt’s studies. If this effect is the same as in Ladefoged and Broadbent’s (1957) study then
The mechanisms behind the spectral compensation effect were hypothesised. A memory process was discussed and central adaptation (such as stimulus specific adaptation) is thought to be involved to keep track of reliability in the signal. This explanation may be preferable to memory as an adaptation process can more clearly account for shifts in test sounds. Memory is not currently known to be a process that causes shifts in the perception of spectra. The work in this section did not examine the importance of spectral variation further, but it was shown that a non-speech precursor that contained this variation was sufficient to bring about the effect. However, it cannot be concluded that Holt’s tests show that the spectral compensation effect is a general auditory process because speech in the test sound may have instigated the compensation seen. Tests without speech are necessary to confirm that the spectral compensation effect is a general auditory process.

Tests without speech: boundary shifts indicating compensation for LTAS were also seen in an entirely non-speech study following Watkins’ paradigm in a study by Stilp et al. (2010). However, there was no confirmation that a central effect was causing the boundary shifts (possibly the enhancement effect was behind this). Therefore, the study does not provide the required evidence to confirm that the spectral compensation effect is a general auditory process. However, it appears to show that at least one of the extrinsic compensation effects that cause shifts in test sounds (the enhancement effect or the spectral compensation effect) occurs in this case. Sjerps et al.’s (2011) tests used sentence-length manipulated speech signals that sounded like non-speech. No compensation was seen where the signal was manipulated so that it was acoustically unlike speech but where speech elements remained in the acoustical signal some small compensation was seen (despite this signal not being heard to be speech-like). This may show that compensation occurs more readily with natural or familiar acoustical signals rather than speech signals per se. This effect may therefore occur with natural non-speech stimuli, but natural stimuli were not tested. It also remains possible that the compensation seen was only caused by a peripheral adaptation mechanism, as crossaural presentation was not tested. A central shift was found in a test without speech by Pitt (1995) but the paradigm, aims of research, and the proposed mechanism of shift were different, preventing a conclusion that the spectral compensation effect
occurs with non-speech. At the very least Stilp et al.’s (2010) study shows the enhancement mechanism causes shifts in a test without speech, Sjerps’ study shows that natural stimuli may be more likely to bring about a central effect and Pitt’s study shows that adaptation to abstract representations may contribute to channel compensation at a non-auditory level. After an examination of the peripheral and central causes of channel compensation at an auditory level (as is done within this thesis), further research into compensation for the channel that occurs through adaptation at an abstract auditory object level could be investigated.

**Two compensation mechanisms working together**: Holt’s results show a lack of effect of the more recent prior spectrum, which raises the question of how the enhancement and spectral compensation processes work together. The fact that both arise from separate mechanisms shows that they may both occur simultaneously but it is not clear how the perceptual effects would interact. Previous research has made little attempt to explicitly explain how both of these effects cause channel compensation individually and there has been no consideration of how the processes might work together to produce this compensation. A brief discussion of how the effects may work to cause channel compensation individually and together is given below.

Many authors investigating the enhancement effect have suggested that enhancement compensates for the channel (Holt and Lotto 2002, Lotto et al. 2003, Summerfield and Assmann 1989, Summerfield et al. 1984). How enhancement causes channel compensation is not explained beyond a description of the effect enhancing change relative to the channel. In running speech (or other programme items) enhancement appears to occur via a compensation for the energy in the prior phoneme. This appears to mean that both the energy that belongs to that prior phoneme (the source) and the contribution of the channel in that phoneme are compensated for when hearing a new phoneme and that the new energy within the new phoneme is perceived to a greater extent. This would cause channel compensation because the channel contribution as well as that of the prior phoneme is reduced in the perception of the current phoneme. If this happens on a phoneme-by-phoneme basis then the channel will always be reduced. However, it is apparent that this enhancement effect may only fully remove the channel if the contribution of the channel is the same with every phoneme. It appears that in practice the whole channel would not be excited with every phoneme, only part of the
channel. Given this, any part of the channel not excited by the prior phoneme but excited by new phoneme will be enhanced along with the new energy relating to the new phoneme. Therefore, compensation for the channel applied on a phoneme-by-phoneme basis will vary and the appropriate compensation may not be applied. Unless the whole channel is excited with the presentation of each phoneme, appropriate channel compensation will not occur. Therefore, a short time course effect such as enhancement may not provide adequate channel compensation. In this case an LTAS mechanism may be more appropriate for removing the channel contribution from every instance of listening.

Watkins (1991) also concludes that enhancement may not be a means of channel compensation for other reasons. He describes the process as not being adequate for channel compensation due to the fact its quick offset means that it would break with gaps in running programme items (though Watkins notes that gaps in running speech may be filled by reverberation) and also because it is not sensitive to direction change or to other cues of channel change, which means that the wrong compensation would be applied after a change in the channel and this would be perceptually disadvantageous. However, these facts are not necessarily a problem because the enhancement is highly responsive. Compensation would begin quickly after a silent gap in the programme item or a channel change. Watkins also notes that the enhancement effect is too small to compensate for the channel. However, this is also not necessarily a disadvantage as compensation is not expected to be complete, just sufficient to remove overlap.

Another question that arises is that, given that enhancement can explain channel compensation by enhancing change, what is the necessity of the spectral compensation effect? It may be that the magnitude of enhancement means that it cannot sufficiently remove the channel, as noted by Watkins. If the enhancement effect reduces the channel relatively by improving the signal to noise ratio of the source, the ‘noise floor’ (the colouration by the channel) may remain the same in absolute terms but the signal is enhanced—i.e. the signal will vary more around this noise floor. If this is the case then, unless the dynamic range of the signal is increased to infinity by enhancement, the noise floor is never completely reduced. If the hearing system has a dynamic range that is too narrow to allow for sufficient enhancement relative to the channel, then the channel effect may not be removed to a large degree. In this case another effect may be
necessary to further remove the channel. The spectral compensation effect can provide a translational shift of the noise floor to a zero noise baseline and so can fully remove the channel allowing the full dynamic range of the hearing system to be used by change. This effect does not appear to rely on increasing the signal. Enhancement’s role may then be one more of removing the effects of the prior phoneme than the channel.

Based on the assessment of all the evidence presented it appears the spectral compensation effect, rather than enhancement, is more likely to be involved in channel compensation. However, the role of enhancement has not been ruled out and a number of authors Summerfield et al. (1987), Watkins (1991), Watkins and Makin (1994) have argued that this can produce channel compensation. Therefore, both may provide an explanation of compensation for the spectral effects of loudspeakers and rooms.

Stilp et al. (2010) and this author observe that the process of compensation for the spectral effects of the channel may be similar to the process of compensation for background noise. It would be interesting to conduct further work to examine the similarity of mechanisms between the two fields. This author also observes that the process may be the same as that involved in compensation for distortion of the temporal envelope of speech caused by the channel (room reverberation) (Beeston et al. 2014, Watkins 2005a). In all cases distortion is caused by a time-invariant additive or subtractive component produced by the channel that is not part of the sound source, and in all cases this reduces the dynamic range of the signal and should be removed to obtain increased sensitivity to signal variation (i.e. to increase the signal to noise ratio). Responses in the Medial OlivoCochlear (MOC) efferent pathway appear to cause compensation for noise and temporal envelope distortion caused by reverberation. This process may also be involved in compensation for spectral colouration. Being a low-level process, and having a quick time course (offset within 500ms (Beeston et al. 2014)), this process may be more suited to explaining enhancement rather the spectral compensation effect. Future research leading on from this thesis could determine whether the auditory model composed by Beeston et al. (2014) can explain shifts in timbre following coloured context sentences, as well compensation for the ‘temporal-smearing’ effects of reverberation on the intelligibility of speech sounds that was found in their study.
Chapter questions

**Question 5.1** asked: a) To what extent does compensation for the spectral effects of the vocal tract occur in speech perception? b) What are the mechanisms behind this? c) What is the time course of this compensation? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

Overshoot and the enhancement effect have been suggested to contribute to channel compensation, including VT compensation. The overshoot studies presented show that a monaural mechanism contributes to overshoot and that the enhancement effect is likely to be behind this. In this section further evidence is presented to show that overshoot and enhancement are the same process. They are both monaural and have a short time course (the central component of the overshoot effect seen in work in this section is unlikely to be part of the overshoot effect; it appears instead to be part of the spectral compensation effect). Overshoot/enhancement may provide general channel compensation, including compensation for the VT via enhancing the changing sound source relative to the channel. If future studies show that there is a short-onset central effect involved in overshoot that is separate to the spectral compensation effect, then this may contribute to overshoot and VT compensation and should be investigated further. However, if this is a manifestation of the spectral compensation effect then this mechanism will not cause compensation for the VT via overshoot, but would still cause compensation for the VT (and other channels) via compensation for the LTAS of the prior sounds.

The work in this section showed a new *acoustic histories* effect, which appears to be a manifestation of the spectral compensation effect, occurring with global non-speech precursors. The two effects are now both considered ‘the spectral compensation effect’. This study confirmed that the spectral compensation effect produces compensation for LTAS: shifts in the perception of test sounds in the contrastive direction to the mean spectrum of a longer precursor occur, rather than in contrast to the immediately adjacent phoneme, or final segment of the precursor. This compensation for the LTAS strongly suggests that this mechanism compensates for the channel, including the VT, as it samples a static spectrum over a period long enough to gain a picture of the channel
response and applies this in compensation. The spectral compensation effect has been shown to directly contribute to compensation for the VT by two studies linking it to Ladefoged and Broadbent’s (1957) work. It is becoming more certain that the spectral compensation effect causes compensation for the VT. However, future research would be desirable to confirm that all studies that purport to show this single central spectral compensation effect.

Further information on the time course of the spectral compensation effect was gained via Holt’s studies. It appears to onset with as little as a phoneme-length precursors (as seen in overshoot studies), but where precursors are longer than the phoneme, compensation is for LTAS of the whole precursor. It is not clear what the window length of the onset is as maximum onset was not observed. The offset of the effect is slower than the enhancement. The effects lasts over at least 1.3s of silence and a 70ms standard tone. The offset window has not been confirmed and may be longer than this (it was as long as 10s in Ladefoged and Broadbent’s (1957) work). The offset of this mechanism means that it does not occur with silence between the precursor and test sound so can be considered time-gap sensitive. The underlying mechanisms behind the spectral compensation effect are suggested to be timbral memory, and/or stimulus specific adaptation occurring at central sites. The MOC noise reduction process may also contribute either to enhancement or spectral compensation. Little new evidence was provided to show that the enhancement/overshoot effect or the spectral compensation effect occurs in an entirely non-speech context as speech perceptions were elicited in most tests. A discussion of the extent to which these effects occur in an entirely non-speech context is saved for Question 5.4.

**Question 5.2** asked: a) To what extent is the phonetic context compensated for? b) What are the mechanisms behind this compensation? c) What is the time course? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

The overshoot effect was investigated further in the studies described in this section. Enhancement was shown to be likely to cause overshoot as overshoot was confirmed to be, at least in part, monaural and has a short time course. A significant monaural effect was seen in overshoot with real-speech studies, which, given enhancement is the process
behind overshoot, suggests that enhancement is not reduced with speech precursors to the extent that it cannot explain overshoot. The exact mechanisms behind this process (e.g. peripheral adaptation) were not investigated further.

Not only does enhancement contribute to overshoot but a central effect may be involved, which was not present in enhancement studies. However, this central effect appears to be a manifestation of the spectral compensation mechanism. Both enhancement and the central component of overshoot may be said to be time-gap sensitive effects, as they recover with a short period of silence, meaning that compensation is not applied where there is a time gap between sounds. The spectral compensation is unlikely to contribute to compensation for the phonetic context as its time course is too long. In running speech compensation for the LTAS of a longer segment will be applied rather than the immediately adjacent phoneme. A discussion of the extent to which these effects occur in an entirely non-speech context is saved for Question 5.4.

**Question 5.3** asked: a) To what extent is the effect of transmission channels other than the vocal tract (e.g. loudspeakers, rooms) compensated for in speech perception? b) What are the mechanisms behind this? c) What is the time course of effects? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

The work in this section largely either involved the investigation of overshoot, which is tied closely to speech processes, or an investigation of the prior spectrum on speech sounds more generally. As discussed before in this review, studies on overshoot may be uncovering a speech specific mechanism because undershoot, the problem that overshoot/enhancement seeks to correct, appears to be a unique speech problem. In which case the overshoot/enhancement process may not produce compensation for channels more generally, when listening to non-speech. However, if enhancement is involved, then this mechanism may produce compensation for other channels as the case is strong that enhancement is a general auditory process. Future work may be considered where enhancement as a cause of overshoot is confirmed and enhancement is confirmed to be a general auditory process. Further, future work should also confirm that the nature of enhancement is appropriate to provide compensation for
the channel. This would show more conclusively that enhancement causes channel compensation. However, enhancement has been frequently discussed as a mechanism of channel compensation, and the evidence so far implies that it can contribute to compensation via enhancing change relative to the channel. Therefore, it is considered to be a mechanism of channel compensation in the remainder of this thesis. Work should also show that enhancement can occur in an entirely non-speech context if it can be a mechanism of channel compensation when listening to any sound. This work is conducted in Chapter 6.

The spectral compensation mechanism appears to be a general auditory process that compensates for LTAS, so may remove effects of any channel. However, entirely non-speech tests are necessary to confirm this mechanism can occur with non-speech sounds. This work is conducted in Chapter 6. Direction sensitivity may still be an issue with this effect when compensating for room reflections. Direction sensitivity may mean that the effect does not compensate for room colouration but the work in this section did not provide further evidence of direction sensitivity, so it would be desirable to investigate this further before its utility in compensating for the room is questioned. So far only one study has suggested that this effect is direction sensitive. In this thesis this mechanism will be considered to have the potential to provide compensation for colouration caused by the room, as well as the loudspeaker. The time course and time-gap sensitivity of these effects were discussed in the answer to Question 5.1 A discussion of the extent to which these effects occur in an entirely non-speech context is saved for Question 5.4.

**Question 5.4** asked: a) To what extent does compensation for transmission channels occur when listening non-speech sounds? b) what are mechanisms are behind this? c) What is the time course of effects? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time gap-sensitive?

Despite the main aim of the section being to determine compensation mechanisms in tests which do not use speech, little information was obtained regarding this. This is because most of the tests purporting to test the speech specificity of these mechanisms used speech in the test sounds. This may instigate a speech mode of perception. The work in this section showed that in overshoot studies identification boundary shifts
occur to a similar extent following non-speech precursors as when speech precursors are used. This is suggestive of a non-speech mechanism behind overshoot, but the use of sine tone transition ‘non-speech’ precursors and speech test sounds may bring about a speech mode of listening, so the non-speech nature of overshoot—including the enhancement and central component of is—is not confirmed via this work. The evidence presented across all 3 literature review sections suggest that enhancement/overshoot is likely to be a general auditory process as its nature suggests a peripheral mechanism is behind this and there is evidence for an enhancement-like effect in masking studies which do not involve speech. However, an entirely non-speech study may be useful to confirm this (see Chapter 6).

Firm conclusions regarding the spectral compensation effect with non-speech listening can also not be made. Holt’s work was the first to test this with a non-speech precursor and shows that the effect occurs. Spectral variation rather than speech in the precursor may be all that is necessary to bring about the effect. A few tests using only non-speech sounds (with spectrally varied precursors and test sounds) were discussed. Three studies appear to show that at least one of these two effects occur with non-speech (either the enhancement effect or spectral compensation or both) but which one could not be confirmed because the central nature of the effect was not adequately tested, the stimuli were unnatural so may not be expected to elicit compensation with any type of programme item, or a central effect was confirmed via crossaural presentation but the study contained too many differences in aims and methodology to confirm that this was the spectral compensation effect. Further work examining these mechanisms with non-speech sounds is carried out as part of this thesis and described in Chapter 6.

5.3.5 Section conclusion

This part of the literature review aimed to contribute to answering Research Question 1b and to contribute towards a process of determining potential mechanisms of the time-gap sensitive component of real-world compensation for loudspeakers and rooms. Specifically, this part of the literature review aimed to determine mechanisms that are known from psychoacoustic research to provide compensation for the spectral effects of the channel when listening to non-speech sounds. These mechanisms might be
potential mechanisms of the time-gap sensitive component of real-world compensation when listening to non-speech (if confirmed to be time-gap sensitive and if Question A-C can be answered for these mechanisms). Chapter Question 5.4 relating to the extent of evidence for compensation for transmission channels in non-speech listening was of particular interest. However, limited information has been provided to answer Chapter Question 5.4 as few studies have been conducted that examine non-speech perception. Almost all studies involved speech in the test, which may have engaged a speech mode of listening. The few studies that have been conducted using entirely non-speech sounds provide inconclusive evidence of compensation with real-world non-speech material and which mechanisms contribute to any compensation seen in those studies is not confirmed.

The literature review (and Step 1a) is now complete. The next section (Section 5.4) assesses whether Step 1 Questions A-C can be answered in the affirmative for any of the mechanisms described in this review with speech and non-speech listening and whether these mechanisms are time-gap sensitive. If this is the case a list of potential mechanisms of the time-gap sensitive competent of real-world compensation for loudspeakers and rooms can be established and Research Question 1b will be answered. It is apparent from the research provided in this section that more work is necessary to examine compensation mechanisms with non-speech and specifically to determine whether some of the mechanisms shown, which may be potential real-world compensation mechanisms (e.g. the enhancement effect and the spectral compensation effect), occur when listening to real-world non-speech items. Therefore, it appears that there may not be sufficient information necessary to answer Step 1 Questions A-C with non-speech and it may not be possible to confirm potential mechanisms of compensation for non-speech listening. Where information necessary to answer Questions A-C is missing Step 1(a)ii set out in Chapter 4, which involves conducting additional psychoacoustic laboratory tests to obtain this, may need to be conducted.
5.4 Establishing potential mechanisms of real-world compensation

An aim of this thesis is to answer Research Question 1b and determine potential mechanisms behind the real-world compensation for loudspeakers and rooms seen in Olive et al.’s (1995) work and the experiments described in Chapter 3. In Chapter 4 it was noted that a research process should be set out to further answer Research Question 1b and determine the mechanisms behind real-world compensation. This research process was described in that chapter. Further, in setting out this research process it was decided that the scope of Research Question 1b should be limited to determining the potential mechanisms behind the time-gap sensitive component of real-world compensation for loudspeakers and rooms. In light of this aim the first step of the research process, Step 1a, was to identify time-gap sensitive mechanisms of compensation for colouration by transmission channels, that are already known about in psychoacoustic research as these might explain real-world compensation. Therefore, a review of the previous psychoacoustic literature was proposed to identify such mechanisms and describe their attributes (step 1a (i)). Once compensation mechanisms were identified via this review it was acknowledged that it would be desirable to confirm which, if any, actually contribute to the time-gap sensitive component of real-world compensation using real-world tests. However, it was also noted that these mechanisms may be considered potential mechanisms of real-world compensation if they are shown to be time-gap sensitive and Step 1 Questions A-C can be answered for each:

**Step 1 questions**

A) For any compensation mechanism, does the mechanism produce a measurable effect that is indicative of real-world compensation and compatible with providing compensation in Olive et al.’s (1995) experiment and in the experiments described in Chapter 3?

B) Is there one or more distinguishing feature of this mechanism (e.g. is it crossaural or does it require a long onset)? This is required to:

- confirm that the mechanism is a unique and distinct mechanism and not a
manifestation of another compensation mechanism and

- separate this compensation mechanism from other compensation mechanisms in tests of elimination in Step 2.

C) Is the time course of the mechanism such that it is compatible with being an explanation of compensation in Olive et al.’s (1995) experiment and the experiments described in Chapter 3?

Additionally, in order to establish that the mechanism in question is a general auditory mechanism that may explain real-world compensation when listening to any sound Questions A-C must be confirmed for that mechanism when listening to speech and non-speech sounds.

The literature review (Step 1 a(i)) has now been completed and this review has provided information on a number of mechanisms that may have the features necessary to account for the time-gap sensitive component of real-world compensation for loudspeakers and rooms. For each mechanism Questions A-C must be addressed and time-gap sensitivity assessed to determine whether it can be considered a potential mechanism of this compensation. When a list of one or more such mechanisms is obtained for speech and non-speech listening then it can be concluded that Research Question 1b is answered and the aim of this thesis is fulfilled.

However, in Chapter 4 it was acknowledged that the literature review may not provide sufficient detail about some of these mechanisms to answer Questions A-C, to determine whether they are time-gap sensitive, or to determine whether they occur when listening to both speech and non-speech. In this case additional experiments may be necessary to provide this information and Questions A-C must be readdressed after these experiments (i.e. step 1a(ii) may need to be conducted). The current section aims to establish whether the information from the literature is sufficient to determine whether the mechanisms in the review are time-gap sensitive and whether Questions A-C can be answered. If so the mechanisms can be entered onto a list of potential mechanisms behind the time-gap sensitive component of real-world compensation. If one or more mechanism can be entered onto this list for speech and non-speech listening, then Research Question 1b is answered. If not then additional experiments must be
5.4 ESTABLISHING POTENTIAL MECHANISMS OF REAL-WORLD COMPENSATION

conducted and these questions addressed again after this information is obtained.

This section is divided into 4 subsections. Whether the mechanisms put forward in the literature review are time-gap sensitive and whether Questions A-C can be answered using evidence from speech tests will be discussed (see Section 5.4.1). Whether the mechanisms put forward in the literature review are time-gap sensitive and whether Questions A-C can be answered using evidence from non-speech tests will be discussed in Section 5.4.2. Conclusions regarding the potential mechanisms behind the non-time-gap sensitive component of compensation are considered in Section 5.4.3 and the extent to which the criteria for determining potential mechanisms of compensation (time-gap sensitivity and answers for Questions A-C in the affirmative) have been met for speech and non-speech listening is summarised in Section 5.4.4 and further work to achieve this is discussed.

5.4.1 Potential compensation mechanisms: Speech

Mechanisms that compensate for the channel when listening to speech are listed in this section and the extent to which Questions A-C can be answered for each is discussed. Their time-gap sensitivity is also discussed.

The literature review showed that a number of mechanisms appear to produce timbral constancy and may be considered mechanisms of compensation for the channel. Some compensation mechanisms involve prior listening and a process that occurs over time, using information beyond the sound that is receiving the compensation. These have been described as extrinsic compensation mechanisms. Others are intrinsic whereby the compensation is immediate and only uses information within the same sound. Both appear to create perceptual compensation for the transmission channels. However, there appear to be broad differences between intrinsic and extrinsic compensation mechanisms which may effect the extent to which these mechanisms contribute to real-world compensation. Therefore, a discussion of the answers to Questions A-C for each of these mechanisms is divided by whether the mechanism is intrinsic or extrinsic.
5.4 ESTABLISHING POTENTIAL MECHANISMS OF REAL-WORLD COMPENSATION

Question A

Question A asks: does the mechanism produce a measurable effect that is indicative of real-world compensation and compatible with providing compensation in Olive et al.’s (1995) experiments and in the experiments described in Chapter 3? Essentially this question asks whether the mechanism has the required nature to provide compensation in the real-world experiments.

Intrinsic processes of compensation in speech perception:

The intrinsic processes described in the literature review can be divided into two types: definitely compensatory and probably compensatory. Probably compensatory mechanisms are:

1) Target perception (single targets) (see Section 5.1.1)

2) Categorisation (see Section 5.1.4)

Common to these mechanisms is the feature of redundancy of information. These processes involve focusing on a limited range of stable cues and ignoring other cues that may be subject to variation caused by the channel. Target perception theory proposes that distortion may be perceived but as long as a couple of necessary key targets are stable, correct identification of a speech sound will occur. Categorisation provides that distortion may be perceived but because multiple target cues are available, some of which are not distorted, correct identification will occur (i.e. target theories require necessary targets and categorisation require sufficient targets). These methods appear to combat certain distortions caused by the channel and aid the identification of speech sounds but it is not clear whether they provide compensation, as distortion to timbre may still be perceived alongside the correct recognition of a sound. It is therefore not clear whether these mechanisms can remove the perception of the channel. Additionally, these mechanisms do not appear to offer compensation where distortions are large enough to push the key targets/enough targets to ambiguous regions where they may cue another vowel. There is no shift in the distorted spectrum of the sound back to its canonical form in this case. Therefore, in many cases distortion may still result in incorrect identification. However, it is assumed that via the perception of
key targets and the correct vowel the impact of the distortion is lessened (see further the discussion in Chapter 7). In which case these mechanisms may be described as compensation mechanisms.

These mechanisms have the nature to contribute to the real-world compensation seen in Olive et al.’s (1995) work and experiments in Chapter 3. If variation caused by the channel is diminished via these mechanisms and stimuli are categorised as the same, or more similar, in spite of different distortions, the contraction of ratings seen with indirect comparisons would occur. Assuming that the channel effects in real-world experiments are not so large as to change the category of the stimuli these mechanisms would provide compensation for the channel in real-world experiments. These mechanisms are therefore confirmed to have a nature that provides channel compensation but their nature means that they may not be able to cause compensation for all variation e.g. larger variation that results in the pushing of spectra into a different region of categorisation (i.e. overlap).

Other means of intrinsic channel compensation are more clearly compensatory. These are named as such because they describe a shift in perception of the sound back to its canonical form, which would not be provided by non-compensatory mechanisms.

3) Categorical perception/PME (see Section 5.1.4)

4) Intrinsic normalisation (see Section 5.1.2)

5) target ratio perception (see Section 5.1.2)

CP results in variation to a target cue or number of targets being reduced in perception and the target/number of targets being shifted back to their canonical location. This may reduce the effect of the channel. However, CP like the target perception and categorisation, does not appear to combat larger distortions. If targets are shifted sufficiently so they cross the boundary to the target for another category, categorisation that results from CP will not shift the targets to the canonical form of the original vowel (or other sound) but the canonical form of another vowel (e.g. the CP boundary for a formant’s locus of origin may be at a 1600 Hz boundary. If distortion to a cue at 1500 Hz is large enough to shift perception across the boundary, categorisation may be
of a canonical sound with an 1700 Hz locus of origin rather than 1500 Hz). CP appears unable to solve the problem of overlap.

In Olive et al.’s (1995) experiments and experiments in Chapter 3, variation caused by the channel may be diminished by this mechanism. Stimuli that are affected by a channel may be categorised and this may result in them being perceived as the same, or similar, reducing perceived variation caused by the channel. However, this is only likely to occur if distortion by the channel is not large enough to push the sounds into a different categorisation region. The mechanism therefore has a nature that means it can explain the channel compensation seen in real-world experiments (providing the colouration is not too large).

Intrinsic normalisation methods involve the perceptual recalibration of any cues distorted by the channel using other cues that occur within the same phoneme, which can be used as reference cues to determine the nature of the distortion (such as F0 and F3 which are thought to describe VT length). Target ratio perception is a similar process whereby the listener perceives targets relative to another or many other targets within the phoneme, i.e. they perceive F2 relative to F1. Both of these methods can be said to involve relative perception rather than absolute perception. This causes a shift of target cues back to a canonical location. These intrinsic mechanisms may reduce overlap if sufficient cues exist for appropriate normalisation. However, sufficient cues may not always exist within a single phoneme.

This normalisation and/or ratio perception may be involved in the channel compensation seen in Olive et al.’s (1995) experiments and experiments in Chapter 3, as they can cause perceptual recalibration and shifts in the perception of a timbre in the opposite direction of the colouration caused by the channel. Sounds within the channel are perceived relative to the channel. This process would bring the perception of a sound coloured by a channel back to a perceptual midpoint. It would bring the perception of sounds coloured by different channels closer together, (a contraction of perception toward a perceptual midpoint), as is seen to occur with indirect comparisons in real-world studies.

**Extrinsic processes of compensation in speech perception:**
Extrinsic processes involve a sample of prior speech being used to normalise speech cues relative to the *spectral range* or the LTAS of the talker, or of the prior phoneme. Extrinsic mechanisms of compensation described in the literature review are:

1. Overshoot (see Section 5.1.3 onwards)
2. VT compensation (Ladefoged and Broadbent 1957) (see Section 5.1.3 onwards)
3. Enhancement (see Section 5.2.1 onwards)
4. The Spectral Compensation Effect (Watkins 1991) (see Section 5.2.2 onwards)
5. The Acoustic Histories Effect (Holt 2006) (see Section 5.3.2 onwards)
6. Low-level (Peripheral) neural adaptation (see Section 5.1.3 onwards)
7. Central neural Adaptation (see Section 5.2.2 onwards)
8. Timbral Memory (see Section 5.2.2 onwards)

These extrinsic mechanisms work in the same way as intrinsic normalisation mechanisms but use a wider range of cues, obtained over the course of longer listening, to more accurately predict the nature of the channel distortion. Like intrinsic mechanisms these mechanisms are regarded as compensatory as they shift the distorted signal back to its canonical form. All appear to provide perception relative to the LTAS of prior sounds. Through sampling the range of cues, the LTAS of the sound can be *calculated* and the listener is able to perceptually remove this via perception that is relative to the LTAS. This recalibration is demonstrated by perceptual shifts in an unfiltered test sound in the opposite direction of filtering to prior sound.

Extrinsic compensation may explain compensation in Olive et al.’s (1995) experiments and experiments in Chapter 3, in the same way as was described for intrinsic normalisation: there is a shifting of perception in the opposite direction to colouration caused by the channel, making a stimulus coloured by different channels sound more like the perceptual midpoint and the same stimuli coloured by different channels more similar. Like intrinsic normalisation methods these methods may be more appropriate for larger distortions that cause overlap as there will be a compensatory shift of all
targets in the opposite direction of the precursor LTAS towards canonical targets. A difference between intrinsic and extrinsic compensation is that the slow response of the extrinsic mechanism provides an enhancement of differences between the channels where there is a quick channel change. The slow response means that old compensation will still be applied to the new channel. Therefore this process both results in compensation for the channel via a reduction in perception that onsets over time, but also an enhancement of channel differences when heard side-by-side. Where there is a time gap between different channels this enhancement will not occur. This may cause the compensation seen in Olive et al.’s (1995) results and the experiments in Chapter 3 because where sounds are presented side-by-side in the direct comparison condition there will be an enhancement of timbral differences between stimuli but this will not occur in the indirect comparison condition where the sounds are separated in time. The extrinsic mechanisms may therefore be described as time-gap sensitive because this compensation built up from hearing one sound will not be applied to a later sound where there is a long enough time gap between the two.

All of the the extrinsic mechanisms provide a measurable effect that is indicative of compensation and may contribute to real-world compensation. However it is doubted whether timbral memory is an extrinsic compensation mechanism. It is extrinsic as memory involves sound outside of the current perceptual target, but memory has not yet been revealed to show an effect that can be regarded as compensatory. As noted previously, memory is a speculative compensation effect. It appears to be involved in compensation but it is not clear that timbral memory causes the necessary shifts in perception. Memory is commonly acknowledged to result in an increase/decrease in certainty (i.e. a decrease/increase in error), rather than causing perceptual shifts. As memory does not appear to cause a compensatory shift, but an increase/decrease in certainty, it does not appear to be able to cause the contraction of perceived variation in loudspeakers and rooms that occurs with indirect comparisons. In terms of the results of the real-world tests, memory would be expected to increase/decrease the noise around the measurement of stimuli being indirectly compared but the mean ratings would be expected remain in the same place. Memory does not appear to explain the translational shift of means towards the midpoint seen in the real-world experiments. As a result, it is concluded that memory does not have the required
nature to explain real-world compensation and it is eliminated from the list of potential real-world compensation mechanisms. However, it is acknowledged that future work should confirm that memory does not cause compensation. A further elaboration on the reason for this exclusion is given in Section 7.2.3 and future work to examine the role of memory in compensation further is also described.

It is also noted that there are aspects of the enhancement effect that mean it may not have the appropriate nature to compensate for the channel. It may not allow for a sampling of the whole channel spectrum because of its short time course. Overshoot suffers from the same problem. Assuming the nature is sufficient to sample the channel spectrum in the necessary way, then it may operate in the same way as the other extrinsic mechanisms. It can cause relative perception and a shift in perception in a way that compensates for the channel. It can also cause enhancement of side-by-side channel differences. Based on the balance of previous evidence suggesting that this is the case, this mechanism is ruled in as a potential mechanism. Others who have investigated the effect have not disputed its nature as a compensation mechanism (but have disputed it on the grounds of its size and longevity of the effect (Watkins 1991)). Therefore, it is concluded that all of the listed mechanisms, except memory, have a nature that is appropriate to explain the compensation for loudspeakers and rooms seen in real-world listening. Question A is answered in the affirmative for these mechanisms.

**Question B**

Question B asks: ‘is there one or more distinguishing feature of this mechanism (e.g. is it crossaural or does it require a long onset)?’ This is required to:

- confirm that the mechanism is a unique and distinct mechanism and not a manifestation of another compensation mechanism; and

- separate this compensation mechanism from other compensation mechanisms in tests of elimination in Step 2.
Essentially this question is asking whether the mechanism in question is a *unique* mechanism and not a manifestation of another. If it is not unique then it may not be considered a potential mechanism of real-world compensation. It would not be possible to distinguish this mechanism to test for its role in compensation in Step 2 tests. The literature review reveals that the mechanisms listed in the previous subsection are not all unique compensation mechanisms but manifestations of the same mechanism. To properly determine the mechanisms of compensation unique mechanisms should be distinguished.

**Extrinsic processes of compensation in speech perception:**

There is strong evidence that overshoot and enhancement effects are a result of a single mechanism, primarily because both are monaural and have a short time course, compensating for the LTAS of the immediately prior phoneme (Holt and Lotto (2002) and Lotto et al. (2003) found evidence of a short time course central component to overshoot. This is unlikely to exist but if so it may contribute to overshoot but can be considered a separate process). They may both have the same neural basis (low-level peripheral adaptation: simple adaptation, adaptation of suppression or adaptation via the MOC efferent pathway). It is concluded that both are the result of a single unique mechanism. This may be termed the *enhancement mechanism* (as enhancement is a basis for overshoot, rather than the other way around) and the term enhancement can be used to describe both the overshoot and enhancement effect. Neural adaptation in the auditory periphery is not considered as a separate mechanism but the cause of the enhancement effect.

The spectral compensation effect, the compensation for VT seen in Ladefoged and Broadbent’s (1957) study, and the acoustic histories effect all appear to be manifestations of the same effect, which is crossaural, has a long time course and is hypothesised to be caused by a central adaptation mechanism (less likely by timbral memory). This effect appears to provide compensation for LTAS of the longer prior context. These effects will therefore all be treated as examples of the *spectral compensation effect* (as Watkins’ work was the first to specifically test a mechanism that compensates for LTAS, his term is used). Timbral memory is not considered further as a compensation mechanism and central adaptation is considered to be the
underlying mechanism behind the spectral compensation effect so is not considered a separate mechanism.

Therefore, from the list of extrinsic mechanisms there is evidence for two unique potential mechanisms of compensation *enhancement* and *spectral compensation*. However, it should be noted that the effects subsumed under these mechanisms may still be found to be separate if distinguishing differences, such as different time courses, or separate origins within the brain are found in any future testing.

**Intrinsic processes of compensation in speech perception:**

Most of the intrinsic mechanisms appear to be unique. Categorisation and CP appear to be unique mechanisms, with one causing sharp identification boundaries that the other does not. Target perception and categorisation appear to be similar but they vary in the number and type of targets that are thought to be necessary to result in correct perception (categorisation requires a number of sufficient targets whereas target perception requires a few necessary targets). These mechanisms will be regarded as separate.

Target perception and target ratio perception appear similar, but perceiving a target ratio is an example of relative perception whereas a using single target is not. Target ratio perception and intrinsic normalisation processes are very similar. Normalisation requires one or two cues that can be used as reference for the adjustment of the other cues, whereas target ratios do not involve particular reference cues but perception that is simply relative to other cues. Conceptually the perception of target ratios and intrinsic normalisation are likely to be the same process. Both will be regarded as instances of intrinsic normalisation.

The intrinsic normalisation appears to be similar to extrinsic normalisation as both result in relative perception. However, intrinsic normalisation is based on less information (that in the phoneme) and occurs over a shorter time course (i.e. *immediately*), whereas extrinsic compensation appears to use more information (that obtained from listening to a longer segment of prior speech) and occurs over a longer time course. They will be regarded as separate mechanisms but it should be noted that the time course of a *short-time course* extrinsic effect (i.e. enhancement) and
intrinsically compensation may not be distinct. The distinction between normalisation using information within the phoneme and from the immediately adjacent phoneme may be blurred. However, one means of distinguishing the two processes is by the fact that neural adaptation in the auditory periphery appears to be a likely cause of short-time course extrinsic processes (its time course is compatible with these) but intrinsic normalisation is frequently described as if it occurs via a cognitive process. It appears that the time course of neural adaptation in the auditory periphery is not quick enough to cause this immediate or near immediate recalibration, another mechanism may be necessary to cause this. Therefore, there may in fact be a distinction due to mechanisms. For the purpose of this review they will be considered distinct mechanisms of compensation.

It can be concluded that there are 2 unique extrinsic mechanisms compensation mechanisms (enhancement and spectral compensation effects) and 4 unique intrinsic mechanisms (CP, target perception, categorisation, intrinsic normalisation). Further research may be necessary to fully establish that the mechanisms set out here are truly unique and that the division of mechanisms is correct. However, this is not necessary for the purpose of this thesis as adequate information is obtained from the literature review to determine unique mechanisms with speech listening. Therefore, it can be concluded that Question B can be answered. Enhancement, spectral compensation, target perception, categorisation and intrinsic normalisation are unique compensation mechanisms.

Because these are unique mechanisms these will all have a unique features that allow the mechanism to be distinguished in tests of elimination in Step 2 of the research process (see Chapter 4). As step 2 will not be conducted this is not discussed further here but see Section 7.2.2 for an example of a test of elimination of the enhancement effect using its unique features (its monaural nature).

Question C

Question C asks: ‘is the time course of the mechanism such that it is compatible with being an explanation of compensation in Olive et al.’s (1995) work and the experiments
Question C aims to establish that the time courses of the mechanisms uncovered by the literature review are such that they can explain real-world compensation. The time course must explain compensation with indirect comparisons and the lack of compensation with direct comparisons in real-world tests. Additionally, for the purpose of explaining the time-gap sensitive component of real-world compensation, whether the mechanisms are time-gap sensitive must be established. This will be discussed alongside Question C. The time course of the compensation effects are not precisely determined in the literature but there is enough information to exclude some from a list of potential real-world compensation mechanisms and potential time-gap sensitive mechanisms.

**Intrinsic processes of compensation in speech perception:**

The intrinsic mechanisms cause compensation that occurs immediately and offsets immediately. Therefore, these are not dependent on a longer time course of listening—neither longer listening periods nor time gaps between stimuli affect the operation of these mechanisms. As a result, these mechanisms would be expected to affect listening in both direct and indirect comparison conditions in real-world tests equally. They would onset immediately when hearing stimuli in the direct comparison condition and the indirect comparison condition. Their offset is also immediate and so they are not time gap sensitive. There appears to be no other reason that these mechanisms would produce larger compensation with the indirect comparisons condition. Therefore, although they are compensation mechanisms, which will be at work in the real-world compensation test, they cannot be considered explanations of the real-world compensation that occurs with indirect comparisons but not with direct and they cannot be considered time-gap sensitive compensation mechanisms.

**Extrinsic processes of compensation in speech perception:**

The extrinsic mechanisms occur over a longer time course and may explain the differences between direct and indirect conditions. The time course of the extrinsic mechanisms can be summarised:
• **Enhancement**

Onsets with 150 ms, Offsets around 500 ms but some reports suggest effect lasts a little as 100 ms or as long as few seconds.

• **Spectral compensation**

Onsets with *sentence-length* stimuli (1-2 seconds) but may begin to onset with only a phoneme’s worth of listening (Holt and Lotto 2002, Lotto et al. 2003, Watkins and Makin 1996b). Maximum onset may require longer than currently tested. The nature of the compensation applied may change as LTAS develops. The effect may be larger or more robust with longer length precursors. Offset or *Recovery* in silence requires longer than enhancement (>500 ms) but this may depend on precursor length. With sentence-length precursors recovery has been shown not to occur with a 1.3 s time gap plus 70 ms standard tone by Holt (2005) and there is evidence that maximum recovery is at 10 s (Ladefoged and Broadbent 1957) with a sentence-length precursor.

The longer onset of these mechanisms means that they might provide more compensation with time spent listening. The enhancement effect is at maximum with only 150 ms so would provide the same amount of channel reduction when listening to directly compared stimuli as when listening to indirectly compared stimuli, as the time for which stimuli are heard in both conditions would allow the effect to reach maximum. This may mean that this effect does not contribute to the difference between direct and indirect comparison conditions in real-world tests. However see further below.

The spectral compensation has a longer onset than enhancement. It may provide increased channel compensation with extended listening. Therefore, there may be more compensation with the longer continuous listening time periods seen in the indirect comparison condition compared to direct comparison condition, as the effect may not have reached maximum with the shorter continuous listening periods seen with direct comparisons. Therefore, this effect may explain the greater compensation with indirect comparisons via increased compensation due to longer continuous listening.

Additionally, both these effects may also cause real-world compensation via their ability
to enhance timbral differences. The slow onset and offset of the effects means that the channel compensation from the prior sound is still applied to a new sound when heard immediately adjacent to each other. This means that when these extrinsic mechanisms are at work, and stimuli are heard side-by-side, there is an enhancement of timbral differences between them. Specifically, in real-world tests there is an enhancement of timbral differences between channels when they are heard side-by-side in the direct comparison condition. Both the enhancement effect and the spectral compensation cause this enhancement. However, in the indirect comparison condition the stimuli are not heard side-by-side but with time gaps between them. Therefore, this enhancement cannot occur. The gaps in Olive et al.’s (1995) real-world test, and those in Chapter 3 appear to be sufficient for both the recovery of the enhancement effect and spectral compensation effect. This lack of enhancement of channel differences with indirect comparisons due to both effects appears to explain some of the reason that timbral differences are perceived to be smaller with indirect comparisons.

These mechanisms or at least the enhancement aspect of these mechanisms can be described as time-gap sensitive: the compensation from one sound is not applied to another with a time gap. In real-world tests this prevents the enhancement of channel differences with indirect comparisons and may result in smaller timbral differences between channels. Therefore, the enhancement effect and the spectral compensation effect have a time course that means that they contribute to real-world compensation. Their time-gap sensitivity means that both mechanisms have the potential to explain the time-gap sensitive component of compensation. However, only the spectral compensation effect appears to have a time course that is appropriate to explain increased compensation with the longer listening seen in real-world tests.

**Potential compensation mechanisms established**

The answers to Questions A-C have been discussed for speech perception and it can be concluded that there are two unique compensation mechanisms that have a nature and time course appropriate to explaining the time-gap sensitive component of compensation seen in Olive et al.’s (1995) work and the experiments in Chapter 3. These mechanisms are the enhancement effect and the spectral compensation effect.
5.4 ESTABLISHING POTENTIAL MECHANISMS OF REAL-WORLD COMPENSATION

These mechanisms offer an enhancement of channel differences when channels are heard side-by-side but not when channels are separated in time. This may explain the reduced channel effect reported with indirect comparisons due to time gaps. The spectral compensation effect may also contribute to the non-time-gap sensitive component of real-world listening via increased compensation for the channel with continuous listening.

Potential mechanisms of real-world compensation have been established based on information that was obtained via speech studies. It is not clear whether any of these effects occur when listening to non-speech. While evidence suggests that these mechanisms might not be speech specific, it is desirable to confirm that these mechanisms can explain compensation when listening to music and other non-speech sounds before they can be regarded as general auditory mechanisms with the potential to explain compensation seen in Olive et al.’s (1995) experiments with music, and compensation for channels generally. The extent to which Questions A-C can be answered for non-speech perception based on the information obtained in the literature review is discussed in the next subsection.

5.4.2 Potential compensation mechanisms: Non-speech

In the literature review it was noted that a number of speech studies show that the mechanisms of speech and non-speech perception may be different. Compensation mechanisms that occur with speech may not occur when listening to non-speech. The literature review has shown that compensation for the spectral effects of the channel has largely only been tested with speech. Therefore, it is difficult to establish whether the mechanisms discussed are potential mechanisms of compensation seen in Olive et al.’s (1995) original study (with music) and other instances of listening to non-speech sounds.

The extent to which evidence from the literature review is sufficient to answer Step 1 Questions A-C for the mechanisms discussed above with non-speech listening is assessed in this subsection. Further, the evidence for time-gap sensitivity of these mechanisms with non-speech is discussed. This section also discusses further work that may be
5.4 ESTABLISHING POTENTIAL MECHANISMS OF REAL-WORLD COMPENSATION

necessary to answer these questions for non-speech listening.

**Question A**

Question A asks: does the mechanism produce a measurable effect that is indicative of real-world compensation and compatible with providing compensation in Olive et al.’s (1995) experiments and in the experiments described in Chapter 3? Essentially this question asks whether the mechanism has the required nature to provide compensation in the real-world experiments.

Some of the intrinsic compensation mechanisms, namely categorisation and categorical perception, have been shown to occur with non-speech. These are likely to contribute to compensation when listening to non-speech, but as described above they do not have the time-course to explain real-world compensation because these mechanisms are intrinsic and there is nothing to suggest that these mechanisms would cause greater compensation with indirect comparisons in real-world listening. Therefore, these mechanisms are not discussed further with regard to real-world compensation in non-speech perception.

Extrinsic mechanisms have been shown to have the appropriate nature (and time-course) to provide real-world compensation with speech. They are compensatory and can account for the shifts of test stimuli towards the perceptual midpoint seen with indirect comparisons in real-world tests. Two unique extrinsic mechanisms were shown to occur with speech: the enhancement effect and spectral compensation effect. However, few tests of these unique extrinsic compensation mechanisms have been conducted with non-speech. Enhancement-like effects were first discovered with non-speech test stimuli in masking experiments. However, masking tests are sufficiently different from the tests that show the effects suggested to be behind channel compensation (i.e. Summerfield’s and Watkins’ enhancement effect tests) and tests that do examine enhancement in a channel compensation context have used speech alongside non-speech sounds: Summerfield’s and Watkins’ tests of enhancement have all elicited speech perceptions and this may be sufficient to engage a speech-mode of perception, so it remains possible that the enhancement effect shows channel compensation does not
occur with non-speech sounds. Similarly, tests concerning the spectral compensation effect have mostly used speech either in the precursor or test sound or both. Only three studies were reported with entirely non-speech sounds. These studies do not adequately test for a central effect so it is not clear whether they provide evidence of enhancement or spectral compensation, use unnatural stimuli that may not be expected to bring about compensation effects, or contain sufficient methodological differences to prevent the conclusion that either enhancement or the spectral compensation occurs when listening to non-speech. Therefore, it cannot be concluded with certainty that the enhancement effect or the spectral compensation effect cause shifts indicative of channel compensation in a non-speech context with natural material. In the process of conducting the literature review, no other compensation mechanisms that occur with non-speech but not speech were shown.

It can be concluded that Question A cannot be answered in the affirmative for non-speech. There is good indication that mechanisms with a nature appropriate to explain real-world compensation (i.e. enhancement and the spectral compensation effect) are general auditory processes, but it is necessary to confirm that they occur in an entirely non-speech test before Question A can be answered in the affirmative. These tests will be conducted as part of this thesis so that Question A can be answered for non-speech perception (see Chapter 6).

If a test that aims to test enhancement and spectral compensation with non-speech show shifts in the perception of test sounds after filtered precursors, that are indicative of compensation (as seen with speech) and these shifts are shown to be monaural/cross-aural/both, then this will be sufficient to conclude that either enhancement/the spectral compensation effect/both are likely to be behind these shift. This is because only extrinsic compensatory mechanisms can cause such shifts (because only these have the slow offset and compensatory nature that would cause such a shift). The only known extrinsic effect that is monaural is the enhancement effect, so any shift in a test sound that occurs in the contrastive direction to the spectrum of a precursor, and is monaural is likely to be caused by enhancement. The only known extrinsic compensation effect that is central is the spectral compensation effect, so any shift in a test sound that occurs in the contrastive direction to the spectrum of a precursor that is central is likely to be this effect. Therefore, if tests with non-speech show such
shifts Step 1 Question A can be flipped: evidence of enhancement and/or spectral compensation shows a mechanism with the nature to produce the time-gap sensitive component of real-world compensation for loudspeakers and rooms (this nature has already been discussed in relation to speech research). Question A can be answered in the affirmative and it will be possible to state that effects with the nature necessary to explain the time-gap sensitive real-world compensation occur in non-speech listening.

**Question B**

Question B asks: ‘is there one or more distinguishing feature of this mechanism (e.g. is it crossaural or does it require a long onset)’? This is required to:

- confirm that the mechanism is a unique and distinct mechanism and not a manifestation of another compensation mechanism and

- separate this compensation mechanism from other compensation mechanisms in tests of elimination in Step 2.

Essentially this question is asking whether the mechanism in question is a *unique* mechanism and not a manifestation of another. The intrinsic mechanisms that occur in speech perception that are unique were discussed in Section 5.4.1. These are not discussed further here as it has been shown that these do not contribute to real-world compensation.

It was noted that only the extrinsic mechanisms (the spectral compensation and the enhancement effect) can offer time-gap sensitive real-world compensation. Because no extrinsic mechanisms have been confirmed to occur with non-speech via research within the literature review, it is not necessary to discuss the uniqueness of these mechanisms in non-speech perception at this stage. However, it was noted in 5.4.2, that further testing will be conducted to confirm that these mechanisms occur with non-speech. If upon further testing it is found that these mechanisms occur to cause shifts in the perception of test sounds in the opposite direction to filtering to the precursor with non-speech, then this question can be addressed.
5.4 ESTABLISHING POTENTIAL MECHANISMS OF REAL-WORLD COMPENSATION

It was noted that if such shifts occur then the mechanisms must be either enhancement or spectral compensation (depending on monaural/cross-aural nature of the effect) as no other compensatory mechanisms are known that can explain such shifts. Therefore, it can be automatically concluded that these are unique mechanisms as they have already been shown to be unique in speech perception. (See Section 5.4.1. There is nothing to suggest that their features, which distinguish them from other mechanisms (such as monaural/central nature and long/short time course) will be very different when testing with non-speech. Therefore, if there is evidence for shifts with non-speech then this implies enhancement and spectral compensation with non-speech and because these have already been shown to be unique, Question B will be automatically answered.

**Question C**

Question C asks: ‘is the time course of the mechanism such that it is compatible with being an explanation of compensation in Olive et al.’s (1995) study and the experiments described in Chapter 3?

The intrinsic mechanisms that occur in speech perception that have been shown to not have a time course appropriate for real-world compensation. This will also be the case with non-speech listening. Therefore these mechanisms are not discussed further here.

However, in Section 5.4.1 it was noted that the extrinsic mechanisms (the spectral compensation and the enhancement effect) have a time course that means they can offer real-world compensation and explain the time-gap sensitive component of this. Because no extrinsic mechanisms were confirmed to occur with non-speech. It is not necessary to determine whether these mechanisms have the appropriate time-course with non-speech listening at this stage. However, it was noted that above that further testing will be conducted to confirm that these mechanisms occur with non-speech. If upon further testing it is found that these mechanisms occur to cause shifts in the perception of test sounds in the opposite direction to filtering to the precursor with non-speech, then this question can be addressed.

It was noted that if such shifts occur then the mechanisms must be either enhancement or spectral compensation (depending on monaural/cross-aural nature of the effect) as
no other compensatory mechanisms are known that can explain such shifts. Therefore it can automatically be concluded that these effects will have the same time course as the enhancement/spectral compensation effects seen with speech. This is because the time course is not expected to vary a great deal between the different programme items, at least not to the extent that these are no longer extrinsic mechanisms that cause compensation in the indirect but not the direct comparison condition in real-world listening tests and the time-gap sensitive component of this. They are expected to maintain this longer than intrinsic time course, with enhancement remaining a shorter time course effect than the spectral compensation effect. Some variation may be expected due to the acoustical differences in programme items but variations should not be large enough to signal a different effect is occurring with non-speech compared to speech. Importantly any variation is not expected to be large enough to change the conclusion that the time course is compatible with explaining real-world compensation and the time-gap sensitive component of this, as was concluded for speech. Question C can be automatically answered in the affirmative upon finding evidence of enhancement and spectral compensation with non-speech.

**Potential compensation mechanisms established**

The information in the literature was not sufficient to answer Step 1 Questions A-C for non-speech. It must be shown that shifts indicative of the enhancement effect and the spectral compensation effect occur in a non-speech test. If there is evidence of a monaural and/or central compensation mechanism for non-speech then it may also be concluded with some certainty that the enhancement effect and/or the spectral compensation effect occur with non-speech and that these are general mechanisms of channel compensation in real-world listening. If evidence for these effects is found then, because these have already been shown to be unique and have a nature and time course compatible with real-world compensation and the time-gap sensitive component of this, Questions B and C can be answered automatically. No differences are expected that will change these answers. These effects can then be concluded to be potential mechanisms of the time-gap sensitive component of compensation for loudspeaker and rooms with non-speech.
Further research is now needed to show that the spectral compensation effect and enhancement occur in a non-speech test. Therefore, tests of these effects with non-speech should be performed (Step 1a(ii) of the research process will be conducted).

5.4.3 Non-time-gap sensitive mechanisms

It was apparent in Experiment 3 that part of the difference between listening over a shorter and longer time course was not due to the time gap between sounds. To fully determine the cause of real-world compensation, the non-time-gap sensitive component should be examined. This thesis has not aimed to determine the non-time-gap sensitive mechanisms of compensation. The literature review did not aim to examine this component of compensation so little was discovered that might explain the non-time-gap sensitive component of compensation. However, compensation via longer continuous listening (and the ratio of listening in an enhanced versus compensated state), onset separation, and reduced attention to timbral change (the listener could not control the changes and changes were less frequent with indirect comparisons), are non-time-gap sensitive factors that might explain reduced perception of timbral differences with indirect compared to direct comparisons in the real-world tests. Future work, beyond the scope of this thesis, would be necessary to establish how non-time-gap sensitive mechanisms (or non-time-gap sensitive components of mechanisms) contribute to real-world compensation.

5.4.4 Section summary and discussion

This section aimed to determine which of the compensation mechanisms described in the literature review are potential mechanisms of the the time-gap sensitive component of compensation for loudspeaker and room colouration in real-world listening. It can be concluded that, for speech listening, the information in the literature is sufficient to answer Step 1 Questions A-C of the research process (see the introduction to this chapter and Chapter 4) and determine time-gap sensitivity for a number of compensation mechanisms. Positive answers to questions A-C and time-gap sensitivity were seen for two mechanisms: the spectral compensation effect and the enhancement
Both of these mechanisms are unique (see Question B in Section 5.4.1). Both have the required nature—they are compensatory and have the ability to produce the shifts in spectra towards the perceptual midpoint, as occurs with indirect comparisons in real-world listening tests, or enhance differences between adjacent sounds (see Question A in Section 5.4.1). Both have the required time course—they do not occur immediately but occur to a greater extent with a longer listening time course so would cause greater compensation with indirect compared to direct comparisons in real-world listening (see Question C in Section 5.4.1). Both are time-gap sensitive—part of the increased compensation due to the longer listening time course with indirect comparisons is because of the time gaps involved and the sensitivity of these mechanisms to time gaps: the mechanisms produce an enhancement of timbral differences between channels that are heard side-by-side but this does not occur when they are separated in time.

It can be concluded that enhancement and spectral compensation appear to explain the role of time-gaps in causing compensation in real-world tests. Therefore these mechanisms can be concluded to be potential mechanisms of the time-gap sensitive component of real-world compensation that occurs when listening to loudspeakers and rooms where speech is the programme item. However, it was acknowledged in the research process set out in Chapter 4, that although these can be regarded as potential mechanisms of this compensation, and Research Question 1b can answered, it would be desirable to proceed to confirming the role of these mechanisms in this compensation using the real-world test paradigm, as these mechanisms may not fully explain this compensation. There might be other, currently unknown, compensation mechanisms that also contribute to the time-gap sensitive competent of compensation in real-world listening and real-world tests would show this. These confirmatory tests are not conducted as part of this thesis but were discussed as part of ‘Step 2’ of the research process discussed in Chapter 4 and are discussed further as future work in Chapter 7.

In order to determine potential mechanisms of the time-gap sensitive component of real-world compensation for loudspeakers and rooms when listening to any sound, it is necessary to identify time-gap sensitive compensation mechanisms for which Questions
A-C can be answered which occur when listening to non-speech. In the current section it was concluded that it is not certain that the compensation mechanisms that are time-gap sensitive and for which Questions A-C can be answered (i.e. enhancement and spectral compensation) occur with real-world non-speech material. Therefore, there was inadequate information to answer Questions A-C (specifically Question A) and determine potential mechanisms of the time-gap sensitive component of real-world compensation when listening to non-speech. It is now therefore necessary to proceed to conducting tests to determine whether these mechanisms occur with non-speech listening (i.e. it is necessary to proceed to Step 1a(ii) of the research process set out in Chapter 4).

Because the extrinsic mechanisms (enhancement and spectral compensation) are the only mechanisms that are time-gap sensitive and for which Questions A-C can be answered (i.e. they are only mechanisms that are unique and have the nature and time course appropriate to explain real-world compensation) only these need to be tested with non-speech. The following experiments will therefore only test for these mechanisms in a non-speech listening context. If these effects occur with non-speech it will be possible to automatically conclude that Questions A-C are answered and they are time-gap sensitive, as these facts are not expected to change with non-speech listening. So it will be possible to conclude that enhancement and spectral compensation are potential mechanisms of the time-gap sensitive component of real-world compensation for colouration caused by loudspeakers and rooms when listening to non-speech, including the music item used in Olive et al.’s (1995) original test. These experiments are conducted in the next chapter.

The literature review did not focus on determining the mechanisms behind the non-time-gap sensitive component of real-world compensation. The literature review pointed to some potential mechanisms, and other mechanisms not mentioned in the review were noted, that may explain this. However, these were not discussed further here. Future work could be conducted to investigate the non-time-gap sensitive component of compensation further.
5.5 Chapter Conclusion

The work in this chapter contributed to answering Research Question 1b by addressing Chapter Questions 5.1 to 5.4. The first part of the literature review primarily addressed Questions 5.1 and 5.2. Research showed compensation for colouration caused by the VT and the phonetic context when listening to speech. Compensation for the VT is not complete but appears to be sufficient to reduce distortion that could otherwise be disruptive to speech perception. Compensation for the phonetic context removes the problem of undershoot and enhances spectral change between adjacent phonemes. Target perception, target ratio perception, intrinsic normalisation, categorical perception, categorisation, extrinsic compensation for the VT and extrinsic compensation for the phonetic context (overshoot) may all be involved in this compensation. The extrinsic mechanisms have a longer time course and are time-gap sensitive and therefore can play a role in the time-gap sensitive component of real-world compensation for loudspeakers and rooms. However, it was noted that these mechanisms might be unique to compensation for VT effects or unique to speech perception. Only categorisation and CP were shown to occur with non-speech listening through the evidence presented in this section.

Section 5.2 described studies that investigated compensation for channels more generally in speech perception (Question 5.3). The enhancement effect and the spectral compensation effect have been hypothesised to be general mechanisms of channel compensation. The enhancement effect provides compensation for the prior spectrum of a short spectrally static (e.g. noise) precursor (the enhancement effect also appears to be present but smaller when real-speech precursors are used instead of static precursors). The onset and recovery of this effect are short. This effect is hypothesised to cause compensation via enhancing spectral change in a source relative to the static spectrum of the channel. This effect is monaural and not direction sensitive. A peripheral mechanism is likely. It was noted that this effect may have the time course and nature to explain overshoot so overshoot may be regarded as a channel compensation process as well as a process of compensation for the phonetic context. However, the time course and monaural nature of overshoot was not sufficiently established in the work described in Section 5.1 to confirm the similarity between
overshoot and enhancement. The spectral compensation effect is central rather than monaural. This effect appears to provide compensation for LTAS of a longer than phoneme-length precursor when listening to real-speech. This effect has been directly shown to explain the compensation for the VT described in Section 5.1. The time course was not clear from the research presented in this section but evidence suggested that the time course is longer than that of the enhancement effect. As this shows the same VT compensation effect described in Section 5.1, then recovery is likely to be as long as 10 seconds. It was observed that a compensation mechanism which provides compensation for the LTAS of the longer prior auditory context would be particularly suited to channel compensation. The filtering applied in the studies in this section was more similar to general channel filtering, showing that these effects may compensate for general channels rather than only occur to remove effects of speech production processes. Further, both enhancement and spectral compensation effects are time-gap sensitive so may provide the time-gap sensitive component of real-world compensation for loudspeakers and rooms.

The enhancement and spectral compensation effects might be potential mechanisms of the time-gap sensitive component of real-world compensation when listening to the music used in Olive et al.’s original experiment and other non-speech sounds but they may be speech specific. Speech specificity (Question 5.4) was investigated in Section 5.3 of the literature review. Little information was gained regarding the working of these mechanisms in non-speech perception because studies elicited speech perceptions within the test, did not conclusively show central rather than peripheral effects, or used unnatural stimuli which may fail to elicit compensation at all, regardless of the programme item. Studies using non-speech precursors and studies using entirely non-speech sounds suggest these mechanisms do occur with non-speech. The apparent peripheral nature of the enhancement effect and evidence of both enhancement and spectral compensation effects occurring with non-speech precursors (e.g. via the acoustic histories effect) support a conclusion that these compensation effects are not speech specific. However, the evidence so far for these mechanisms working with real-world non-speech sounds is not conclusive.

Although the non-speech studies presented in section 5.3 did not provide conclusive evidence of the discussed compensation mechanisms with non-speech perception, the
work can at least be concluded to provide further information on the existence and
time course of these mechanisms with speech. Overshoot studies with non-speech
sine-tone precursors confirmed that enhancement and overshoot effects are the result
of the same process. Both effects have now been shown to be monaural and it
is confirmed that they have a very similar time course. Further, it was directly
confirmed that the effect elicited by non-speech acoustic history precursors is the
same as the spectral compensation effect, shown in Watkins’ studies with speech.
Therefore, data from the acoustic histories studies can be used to (a) support the
existence of an LTAS compensation mechanism which provides VT compensation (and
probably other compensation for other channels), i.e. support the existence of the
spectral compensation effect; and (b) provide extra information on the time course
of this spectral compensation effect. Acoustic histories studies show that the spectral
compensation effect lasts beyond a 1.3 second silent gap and show more directly that
more than a phoneme’s worth of prior spectrum is used in compensation, where this is
available.

Section 5.4 assessed whether Step 1 Questions A-C can be answered in the affirmative
for any of the mechanisms described in this review with speech and non-speech
listening and whether these mechanisms are time-gap sensitive. Those that meet
these criteria were to be considered potential mechanisms of the time-gap sensitive
component of real-world compensation for loudspeakers and rooms. Data from the
literature was sufficient to answer Question A-C and determine time-gap sensitivity
for a number of mechanisms for speech listening. Questions A-C could be answered
affirmatively for the spectral compensation effect and the enhancement effect as these
are unique, have the ability to produce compensatory shifts as seen in real-world
listening tests, and have the required extrinsic time course which means they may
occur to a greater extent with longer listening and cause enhancement of adjacent
spectra. Further, they are time-gap sensitive and can explain compensation that
results from a separation of sounds. Questions A-C could not be answered in the
affirmative for the other mechanisms mentioned in the review. Therefore, only the
enhancement and spectral compensation effects have the potential to contribute to
the time-gap sensitive component of compensation in real-world tests. More work is
necessary to examine compensation mechanisms with non-speech as not enough data
exists to address Questions A-C and time gap sensitivity when listening to non-speech. Additional tests to determine whether the compensation mechanisms described in the review occur with non-speech listening were proposed (i.e. the research will proceed to Step 1a(ii) of the research process). Because the extrinsic mechanisms (enhancement and spectral compensation) are the only mechanisms that are time-gap sensitive and for which Questions A-C could be answered in speech listening and because no other mechanisms that occur in non-speech perception were found, only these need to be tested with non-speech. If these effects occur with non-speech it will be possible to automatically conclude that they are potential mechanisms of the time-gap sensitive component of real-world compensation for colouration caused by loudspeakers and rooms when listening to non-speech, including the music item used in Olive et al.’s (1995) original test.
Chapter 6

Experiments to test time-gap sensitive mechanisms with non-speech

In Chapter 5 a list of potential time-gap sensitive mechanisms of real-world compensation for loudspeakers and rooms was compiled via a literature review. Of the mechanisms described in the literature review only two unique mechanisms were shown to have the required nature and time course to provide this compensation. These mechanisms are the enhancement effect and the spectral compensation effect. It was concluded that there is sufficient evidence to confirm that these mechanisms have the potential to cause this compensation when listening to speech but, only if the mechanisms also occur when listening to non-speech can they be regarded as general mechanisms of compensation and confirmed as potential mechanisms of the compensation for loudspeakers and rooms seen in Olive et al.’s (1995) original experiment with music. Therefore, tests should be conducted to confirm that these mechanisms occur with real-world non-speech stimuli (i.e. Step 1a(ii) of the research process will be conducted for non-speech listening. See Chapter 4).

The work described in the present chapter describes experiments to test the enhancement effect and the spectral compensation effect with non-speech. These experiments use the same psychoacoustic laboratory paradigm as used in the previous
tests of these mechanisms described in the literature review. If these mechanisms are shown with non-speech in these laboratory tests then, as discussed in the previous chapter (see Section 5.4), Step 1 Questions A-C of the research process (see below and Chapter 4) can automatically be answered and time-gap sensitivity confirmed. These mechanisms can be regarded as potential mechanisms of the time-gap sensitive component of compensation for loudspeakers and rooms in real-world listening with non-speech and research Questions 1b will have been answered for non-speech listening.

Step 1 Questions:

A) For any compensation mechanism, does the mechanism produce a measurable effect that is indicative of compensation and compatible with providing the real-world compensation seen in Olive et al.’s (1995) experiment and in the experiments described in Chapter 3?

B) Is there one or more distinguishing feature of this mechanism (e.g. it is crossaural or does it require a long onset)? This is required to:

- confirm that the mechanism is a unique and distinct mechanism and not a manifestation of another compensation mechanism and

- separate this compensation mechanism from other compensation mechanisms in tests of elimination in Step 2.

C) Is the time course of the mechanism such that it is compatible with being an explanation of the real-world compensation seen in Olive et al.’s (1995) experiment and in the experiments described in Chapter 3?

Experiments 4 and 5 aim to confirm spectral compensation and enhancement with non-speech. However, before these experiments are presented pilot experiments are described. These pilot experiments aim to confirm that the stimuli to be used in Experiments 4 and 5 will be appropriate for examining enhancement and spectral compensation. They also aim to confirm that the stimuli would be suitable for use in future work that may further examine these compensation mechanisms, using the same laboratory paradigm (see Chapter 7). It is necessary to confirm that not only are the stimuli suitable for Experiments 4 and 5 but also future tests as it will be highly
advantageous to be able make direct comparisons between the results of Experiments 4 and 5 and any future tests. This can only be done if the same stimuli and comparable conditions are used across all tests.

A general rationale for Experiments 4 and 5 is described in Section 6.1, a general hypothesis is described in Section 6.2 and a general method is described in Section 6.3. Pilot experiments ensuring the suitability of the material for Experiments 4 and 5 and future work are then described in Section 6.4 and Section 6.5. Experiment 4 is described in Section 6.6. The main aim of Experiment 4 is to show that there is evidence of compensation when listening to non-speech stimuli. The main aim of Experiment 5 (see Section 6.7) is to determine whether there is any evidence that the specific mechanisms shown to be potential time-gap sensitive mechanisms of real-world compensation with speech—the enhancement effect and the spectral compensation effect—cause the compensation seen in Experiment 4. If it is shown that these speech mechanisms occur with non-speech then they may be regarded general auditory mechanisms of compensation for loudspeakers and rooms. A discussion of the extent to which the evidence suggests that the spectral compensation effect and the enhancement effect occur in non-speech perception and the extent to which these can be regarded as potential mechanisms of the time-gap sensitive component of compensation for loudspeakers and rooms when listening to non-speech (i.e. the extent to which Step 1 Questions A-C are answered and time-gap sensitivity is established for non-speech) is presented in the summary and discussion section of Experiment 5 (see Section 6.7.6) and the summary and discussion section of this chapter (see Section 6.8). The extent to which Research Question 1b can be answered using the information presented in this chapter is discussed in the chapter conclusion (see Section 6.9) and in Chapter 7.

The following chapter questions are posed to guide the research in this chapter. These will be answered in the summary and discussion section of this chapter (see Section 6.8):

Q 6.1) To what extent does the work in this chapter show that any of the specific compensation mechanisms cause compensation in non-speech perception?

Q 6.2) Exactly which of the compensation mechanisms result in this compensation
with non-speech?

Q 6.3) Are any of these mechanisms potential mechanisms of compensation for loudspeakers and rooms in real-world listening to non-speech?

6.1 General Rationale

This section describes a general rationale and method that applies to Experiments 4 and 5 (and their pilots) and tests that use the laboratory paradigm, which may be conducted as part of future work to further examine mechanisms involved real-world compensation and complete the research process set out in Chapter 4 (see Chapter 7 for details of future experiments).

In Section 5.4 only the extrinsic compensation mechanisms (the enhancement effect and the spectral compensation effect), were shown to have the potential to cause the time-gap sensitive component of real-world compensation. Therefore, only these are investigated further as potential mechanisms of the time-gap sensitive component of real-world compensation with non-speech. The psychoacoustic laboratory paradigm that is traditionally used to measure these extrinsic compensation mechanisms with speech will be adopted to test for these mechanisms. The rationale for testing the potential role of these mechanism in real-world compensation using the laboratory paradigm before testing their actual role in real-world compensation using real-world paradigm was discussed in Chapter 4. Essentially, the rationale is to confirm that the spectral compensation effect and enhancement effects are potential mechanisms of real-world compensation with non-speech using simple laboratory tests before any future work is conducted to test the role of these mechanisms in real-world listening to non-speech, using the more complex real-world paradigm. Further, these tests offer a direct parallel to the existing research with speech and therefore comparison with speech research can be made more easily with these laboratory tests compared to real-world tests.

However, it was also noted in Chapter 4 that tests using the laboratory paradigm can confirm the working of the extrinsic compensation mechanisms with non-speech,
but the this paradigm measures compensation over relatively short time courses, milliseconds to a few seconds, rather than over the longer listening time courses seen in the real world. It also measures compensation in a more isolated way than the factorial real-world tests. Therefore, having concluded that these mechanisms occur with non-speech in laboratory tests, the extent to which these mechanisms actually contribute to the compensation seen in real-world tests will remain uncertain. The results of the laboratory tests will only show that the enhancement and spectral compensation effects are potential sources of real-world compensation. As discussed in Chapter 4, further tests would be desirable to confirm the actual role of these mechanisms in real-world compensation with non-speech. Such tests may be conducted as part of future work and are described in Section 7.2.

In the current tests compensation for the spectrum of the channel when listening to non-speech, after hearing a precursor consisting of the same sound filtered by a channel, will be measured by examining the extent to which the timbre of the test sound shifts in the contrastive direction to the filtering of the precursor. This method closely follows the traditional psychoacoustic laboratory paradigm, as used in previous extrinsic compensation work (Holt 2006, Stilp et al. 2010, Summerfield et al. 1984, Watkins 1991).

In such laboratory experiments shifts in the perceived timbre of a test sound, in a contrastive direction to the precursor filtering, indicate extrinsic compensation for the transmission channel because shifts show that perception is relative to the prior spectral context and may show compensation to the spectral context and therefore compensation for the channel. The exact underlying mechanisms behind these shifts are not well understood. Such shifts are caused by extrinsic compensation mechanisms and the spectral compensation effect and the enhancement effect (being the only known existing compensation mechanisms) are likely to be behind this compensation. At least for the enhancement effect, a neural adaptation mechanism is a likely underlying mechanism. Summerfield et al. (1984) aimed to explain how shifts in timbre seen in these studies indicate compensation using Smith’s (1979) peripheral simple neural adaptation model. This was described briefly in Section 5.2.1 and is explained further here.
When a precursor sound is presented frequency sensitive regions in the ear and/or brain that are stimulated undergo a gradual reduction (adaptation) of neural firing during the presentation of the sound (to use the terminology of Smith (1979) adaptation occurs to the pedestal). Recovery of this occurs with a silent gap following the precursor but if another test sound (or an increment, (Smith 1979)) is introduced immediately, or soon after the precursor, so that recovery is not complete then the response to previously stimulated energy regions is reduced relative to the new energy introduced by the test sound (adaptation of suppression may also occur, further increasing the new energy). This results in the perceived spectrum of the test sound being pushed further from that of the precursor compared to the situation where there is no precursor. Without a precursor the listener will hear the sound as described by its frequency spectrum. This adaptation has the effect of making timbral differences that are heard side-by-side larger or enhanced and it also diminishes the effect of the static spectrum. This process may describe shifts caused by the enhancement effect and/or the spectral compensation effect, if either or both mechanisms have a neural adaptation basis.

Although Summerfield describes how shifts indicate compensation using a neural adaptation explanation, neural adaptation is not the only process by which shifts indicative of compensation can occur. They may be caused by any cognitive mechanism that causes relative perception. Shifts via relative perception can be explained as follows: after a bright precursor is heard, a following mid-frequency sound may be heard to be dark, relative to the bright context that is established with the precursor. Therefore, shifts in perceived spectrum in the opposite direction to that of the comparison sound show compensation, but do not necessarily show whether a neural or cognitive mechanism is involved; they also do not necessarily show that the enhancement or spectral compensation effects are working to cause these shifts, but as these appear to be the only compensation mechanisms capable of producing such shifts as was concluded by the literature review, this is considered highly likely. Therefore, although it remains possible that other mechanisms exist that can cause such shifts via neural adaptation or cognitive means, which were not discussed in the literature review. For the purpose of this thesis shifts indicative of compensation will be concluded to be evidence of the enhancement or spectral compensation mechanisms.
6.2 General hypothesis

The main hypothesis of Experiment 4 is that in a non-speech test any spectral magnitude region reduced in level by filtering to a precursor sound will be enhanced in the following comparison sound, causing a shift in perception of the spectrum of a comparison sound in the opposite direction to the filtering in the precursor. Such a shift is indicative of compensation for the channel by the enhancement and/or spectral compensation effect.

The main hypothesis of Experiment 5 is that in a non-speech test shifts indicative of compensation will be monaural, crossaural or both. This provides evidence for the enhancement effect, the spectral compensation effect, or both in producing compensation. It is expected that a crossaural shift, indicative of a spectral compensation effect, will occur with music as this contains spectral variation. It is expected that any compensation occurring with noise will only be caused by the monaural enhancement effect.

6.3 General method

Compensation will be measured using the psychoacoustic laboratory paradigm as used in studies of extrinsic compensation in Chapter 5. The specific method used in the current tests is closest to that used by Watkins (1991). The current experiments involve presenting the listener with a filtered precursor sound (e.g. a short excerpt of music) and measuring shifts in the perceived timbre of a test sound (e.g. the same excerpt of music, but unfiltered) or a range of test sounds (e.g. a continuum of test sounds as used in Watkins’ (1991) work).

Due to the use of non-speech test sounds, the perception of which is less easily identified and labelled compared to that of speech test sounds, a standard reference sound will be used to assist in timbre labelling. The use of a standard means that the method deviates slightly from that of Watkins (1991) and the other tests using the laboratory paradigm. In speech studies a standard is not necessary because speech timbres (e.g. vowels) are more familiar and can be labelled more easily when presented in isolation.
Musical timbres are less familiar and it may be more difficult to simply label these consistently without a reference. The use of a standard provides a constant reference and allows the more difficult labelling task to become an easier comparison task.

Specifically, in the experiments described in this chapter, listeners were asked to judge the timbre of test sounds as being ‘bright’ or ‘dark’ after hearing differently filtered precursors. Shifts in their tendency to label these test sounds as bright or dark will be measured. As it is likely that listeners will find the task of labelling sounds bright or dark in isolation more difficult than labelling vowels in isolation (or the common instrument sounds used in Stilp et al.’s (2010) test), the listener will be asked to discriminate the sounds as ‘brighter’ or ‘darker’ than a middle brightness ‘standard’. In the current test the precursor is inserted between a standard and test sound which must be compared for brightness. As in Watkins’ (1991) test, shifts in the perception of test sounds are expected but these will be observed via changes in the labelling of the test sounds as compared to the fixed standard.

To instigate channel compensation the precursor presented before the comparison sound is filtered. As for the test sounds, the precursors are either filtered to be ‘bright’ or ‘dark’. It is expected that the listener will make more ‘brighter than the standard’ responses to test sounds following a precursor filtered to be dark compared to a precursor filtered to be bright and, similarly, more ‘darker than the standard’ responses to a test sound following a precursor filtered to be bright compared to a precursor filtered to be dark. To reflect the fact that judgements of timbre of the test sound are in comparison to a standard, the term ‘comparison sound’ is used instead of ‘test sound’ hereafter.

### 6.3.1 The brightness continuum

As in Watkins’ (1991) work perceptual shifts in response to filtered precursors will be measured using a range of comparison sounds, rather than just a single one. The main reason for using a range is to prevent the timbre of the comparison sound becoming predictable on each trial. However, testing a range of comparison sounds is also useful for testing the effects of compensation at different levels of sensitivity to
timbral differences between the standard and the precursor (see Section 6.3.3).

A continuum of comparison sounds is formed whereby each has a different \textit{brightness}. A comparison sound continuum is formed by creating a number of sounds that vary in their LTAS and are intended to create perceptions that vary from \textit{bright} to \textit{dark}. This is done by taking a short excerpt of the programme item and boosting or cutting High Frequency (HF) content above 1kHz by steps of 1dB, or less, for the range $\pm 3$ dB. Brightness is related to the proportion of HF content within a sound (Schubert and Wolfe 2006), so this continuum was expected to result in a number of comparison sounds that vary perceptually in ‘brightness’. This continuum is referred to as the brightness continuum.

A standard sound is composed consisting of the same excerpt, but with 0dB filtering above 1kHz applied to it (i.e. it is the same the midpoint of the brightness continuum). During the test, sounds from the brightness continuum are paired with the standard. This creates a number of standard/comparison (S/C) pairs that must be discriminated during the task. For each pair the listener must determine whether the comparison sound is ‘brighter’ or ‘darker’ than the standard. When presented with all S/C pairs from the continuum, but without a precursor between the pair, listeners are expected to make responses that vary symmetrically around the standard from ‘darker’ than the standard to ‘brighter’ than the standard. Listeners are expected to be very sensitive to large differences (e.g. $\pm 3$dB S/C pairs); they are expected to discriminate these with 100% or near 100% accuracy (\textit{perfect} or \textit{high} accuracy). Listeners are expected to be less sensitive to the difference with smaller S/C intervals. Comparison sounds at the continuum midpoint are expected to elicit responses of ‘brighter than the standard’ 50% of the time (chance or \textit{low} accuracy). With a large number of trials performance accuracy should produce a smooth sigmoid function, as is commonly seen in similar psychophysical discrimination tasks. In this case the smooth psychometric function is symmetrical and located in the centre of the x axis.

To test compensation, filtered precursors are inserted between the standard and comparison sound. Two types of precursors are used: a ‘dark’ precursor (the same stimulus as in the -3dB continuum endpoint) or ‘bright’ precursor (the +3dB continuum endpoint). These are expected to shift the perception of comparison sound
timbre in the opposite direction to the filtering of the precursor. More ‘brighter than the
standard’ responses are expected after a dark precursor (in this case the continuum
should shift along the x axis to the left) and fewer ‘brighter than the standard’ responses
are expected after a bright precursor (in this case the continuum is shifted to to the
right of the x axis).

6.3.2 Measuring shifts using the category boundary

Shifts caused by precursors can be measured across the whole continuum of comparison
sounds by examining shifts of the whole psychometric function with bright and dark
precursors. However, instead of measuring the shift of the whole function shifts can
be measured at specific points along the function, as was done in experiments by
in the Category Boundary (CB). The CB represents the point at which responses
in a discrimination task change. In the current task it represents the point on the
comparison sound continuum that responses change from ‘brighter’ to ‘darker’ than
the standard. For example, every time a listener performs the task they must compare
a comparison sound to the standard. They must make a binary decision as to whether
the current comparison sound is brighter or darker than the standard. When the
listener discriminates all sounds from the continuum against the standard, there should
be a distinct point along the continuum at which their labelling decision changes from
‘brighter’ to ‘darker’ than the standard. Where performance is accurate this should be
at the continuum midpoint where the sounds do in fact change from darker to brighter
than the standard.

When the listener only judges each sound from the continuum once, the CB is the
point where labelling changes sharply from ‘darker’ to ‘brighter’. However, it is usual
in listening tests of this type for multiple measurements to be obtained for the same
continuum (to get a more accurate measure). In this case the CB would be defined as
the point at which labelling changes from 100% ‘brighter’ to 100% ‘darker’ responses.
However, the listener is rarely perfectly sensitive and consistent, a sharp decision point
is unlikely to occur in practice. Where performance accuracy is less consistent and the
listener is less sensitive, discrimination will be at or near 100% accuracy at endpoints
and at 50% accuracy (low accuracy) at the midpoint. In this case the location of the CB corresponds to the point on the continuum where 50% accuracy is obtained, as this can be regarded as the point at which labelling changes: above this point the listener tends to regard the comparison stimuli as being brighter than the standard and below this they tend regard them as darker than the standard. Obtaining the CB involves repeated measurement for the same continuum points which may be obtained across a single listener or using a number of different listeners.

Watkins (1991) measures compensation using shifts in the location of the CB, with differently filtered precursors. Where there is no effect of the precursor the CB point is expected to be located at the point on the continuum, where the standard is located (e.g. with a 0 dB standard the CB would be located at 0 dB—the midpoint of the continuum). Where shifts occur due to compensation the CB will correspond to a point on the continuum somewhere other than where the standard is located. For example, if perception shifts toward hearing the comparison sounds as more bright the CB will be < 0 dB (if a 0 dB standard is used) This CB measure will be used to measure compensation in the current experiments. It will be denoted CBdB(50%) to show that it is a measurement of the location on the continuum (measured in dB), where 50% ‘brighter’ than the standard responses occur. In Watkins’ experiments it was possible to calculate a confidence interval around this point (because there were repeated ratings by the same listener). In the current experiments this is not possible due to a lack of repeated ratings by the same listener. Therefore, another equivalent measure of compensation is also used for which confidence intervals can be calculated. This is referred to as CB50%(0 dB) and it is measure of the percentage of ‘brighter than the standard’ responses at the 0 dB continuum point. Where shifts due to compensation occur, the 0 dB point will correspond to a point on the y axis of the psychometric function plot other than 50%. This measure is the reciprocal of the CBdB(50%) measure.

Both CBdB(50%) and CB50%(0 dB) measure compensation at the midpoint of the continuum. This is where maximum compensation is expected (see Section 6.3.3), but this provides a limited view of compensation. It is possible to use CB measures to measure compensation at other points on the continuum. For example, the CBdB(25%) and CBdB(75%) can be examined. These measures are frequently used
in psychophysics to find the Just Noticeable Difference (JND) between a standard and sounds on a continuum that vary around this. Asymmetry of the JND values between the two sides of the continuum in tests with precursors indicates compensation (e.g. a CBdB(25%) point located at -2 dB and a CBdB(75%) located at +1 dB indicates greater sensitivity to perceiving stimuli that are brighter than the standard and a bias towards hearing the stimuli as more bright).

The most complete measure of compensation can be obtained by measuring the shift at every continuum point and aggregating this across the whole continuum. This measure is a measure of overall compensation in response to the precursor (CB%(Overall) and CBdB(Overall)). This measure results in the most comprehensive measure of compensation and provides increased statistical power as all data points are used in this analysis, but it may be diluted by lack of shifts at various continuum points, such as endpoints (as is discussed in Section 6.3.3). As it is desirable to have a complete picture of compensation at different points on the continuum (namely to measure compensation at different sensitivities, see further Section 6.3.3), the whole range of measures will be used to describe compensation in the described experiments in this chapter. Table 6.1 summarises all the measures.

6.3.3 Sensitivity and contrast

The issues of sensitivity and contrast may be important to measuring compensation using the current paradigm and should be considered before the tests are conducted. Previous real-world tests discussed in Chapter 3 described compensation as a process by which sensitivity to differences between channels is reduced. Loudspeaker and room ratings were made and whether listeners were less sensitive to timbral differences between channels with indirect compared to direct comparisons was measured to determine compensation. In the current tests perceptual shifts, rather than decreases in channel sensitivity, are used to measure compensation. It would be possible to use sensitivity to measure compensation using the current paradigm because shifts caused by precursors also enhance discrimination accuracy by boosting the difference between
### Table 6.1: Category boundary measures of compensation.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CB% (overall)</td>
<td>The proportion of ‘bright’ responses across the overall continuum: should be 50% without precursors. Compensation measured may be diluted by lack of shift at some continuum points. This measure is termed ‘CB%’ because the CB is given with respect to the percentage of bright responses.</td>
</tr>
<tr>
<td>CB% (0 dB)</td>
<td>The proportion of ‘bright’ responses at the continuum midpoint (0 dB): should be 50% if shifts do not occur, as here listeners should be uncertain of whether the stimulus is bright or dark and respond at chance.</td>
</tr>
<tr>
<td>CBdB (50%)</td>
<td>This is the reciprocal of the CB% (0 dB) measure and is a traditional measure of the CB used by Watkins (1991). This measures shifts at the continuum midpoint. The expected value is 0 dB gain (the continuum midpoint). With precursors the CBdB 50% point is expected to move along the continuum (x axis).</td>
</tr>
<tr>
<td>CBdB (25%, 75%)</td>
<td>The locations of the 25% and 75% ‘brighter’ response points on the continuum (x axis): equivalent to the JND in traditional psychophysics tasks. Shifts caused by precursors will result in the JNDs being placed non-symmetrically about the midpoint. Increased sensitivity will result in these values being placed closer to 0 dB</td>
</tr>
<tr>
<td>CB% (continuum 1/4 point and continuum 3/4 point)</td>
<td>This is the reciprocal of the JND measures.</td>
</tr>
</tbody>
</table>


the precursor and comparison sound where precursor filtering acts in favour of the
difference (i.e. it is facilitative, boosting the difference) and reduces discrimination
accuracy and where it does not act in favour (it is non-facilitative). Therefore, changes
in sensitivity before and after appropriate compensation could be measured with
this paradigm. However, instead of measuring the effect of precursors on sensitivity,
compensation is more simply measured by examining the extent of shifts in perception
in a given direction. This method of measurement is preferable because a) shifts, rather
than sensitivity, are used to measure compensation in the studies discussed in the
1991) and b) measuring shifts is a more direct measure of compensation. Measuring
changes in sensitivity shows compensation but also changes in sensitivity caused by
other factors (e.g. timbral memory).

Although sensitivity is not used as a measure of compensation it is still an issue in the
current experiments. Specifically sensitivity at baseline—the ease with which listeners
can discriminate between the standard and comparison sounds before the insertion
of precursors—is an issue, as this may affect the extent to which shifts caused by
precursors occur or can be measured. In Experiment 1 (see Section 3.1) it was observed
that the extent to which compensation can be measured depends on the existing
spectral differences between stimuli before compensation. In that experiment too small
differences between loudspeakers and rooms (and therefore too low sensitivity) meant
that the perceived differences between them were not significant and could not decrease
further with compensation: there was a floor effect that prevented compensation being
measured. In the current experiments where compensation is measured via shifts,
ceiling effects, rather than floor effects, may occur. With too large differences between
the standard and comparison sound (too high baseline sensitivity) ceiling performance
is expected whereby a shift that is facilitative (a shift in the same direction as the
difference being measured) cannot be observed because where sensitivity is already
good shifts may not have a material effect on the listener’s labelling of the sound as
‘bright’ or ‘dark’. In the case of potential ceiling effects shifts could be measured by
only looking at precursors which cause shifts in the opposite direction of the difference
(non-facilitative). Such shifts are not suppressed by ceiling performance but another
issue may affect the ability to observe these shifts. Where there is a large difference
between the standard and comparison and sensitivity is high, a shift in the opposite
direction may have to be particularly strong to result in a shift in labelling and shifts
may be less apparent (at least over short-term statistics).

Therefore, it should be acknowledged that compensation effects may not be measurable
where baseline sensitivity is high. It is thus desirable to test compensation at
a variety of different sensitivities to examine this issue further. Because of the
mentioned potential issues relating to ceiling performance (with facilitative shifts) and
less apparent shifts (with non-facilitative shifts) at high baseline sensitivity points,
compensation is expected to be more apparent in the middle of the comparison sound
continuum where baseline sensitivity is lower and both of these issues are less of a
problem. Using a range of comparison sounds should mean that compensation can be
tested at a range of baseline sensitivities between the standard and the comparison.
The range of comparison sounds should be chosen so it encompasses comparison stimuli
that will be discriminated from the standard with high and low baseline sensitivity to
allow for measurement of compensation at the full range of sensitivities.

Another issue relating to the fact that the extent of the shift may vary with baseline
sensitivity relates to comparison of shifts across different experiments (or different
conditions within an experiment). If sensitivity affects the extent to which shifts may
occur it may be desirable to compare compensation at the equivalent sensitivity points
across experiments, rather than simply at equivalent points on the continuum, as was
done in Watkins’ work. In order to make this comparison across tests, it will be
necessary to ensure that equivalent baseline sensitivity points are obtained across all
conditions and all experiments (see Section 6.4.6).

As well as sensitivity between the standard and comparison sound, the issues of contrast
between the precursor and comparison sound may be relevant to the extent to which
compensation can occur, or be measured (see Section 5.2.1). The current tests do not
allow for an investigation of contrast separate to sensitivity as to some extent contrast
co-varies with sensitivity in the current experiments. However, some of the effects of
sensitivity discussed may be confounded with effects of contrast, therefore, it might be
interesting to examine this issue in future tests.
6.4 Pilot 1: Music

This section describes the first pilot experiment. This pilot examines the perception of the brightness continuum that will be used to measure compensation effects in Experiments 4 and 5 in this chapter and in future tests using the laboratory paradigm that might be conducted as part of the work following this thesis (see Chapter 7). The main aim of this pilot is to determine validity, sensitivity and bias when judging sounds from the brightness continuum, as these issues may affect the extent that compensation can be measured in experiments 4 and 5 (and any future tests). In the current pilot listeners were asked to discriminate between sounds from the continuum against the standard by labelling them as ‘brighter’ or ‘darker’ than the standard. No precursors were inserted as the aim was to measure the perception of the continuum rather than the effects of compensation on this. The aims of this test were as follows.

6.4.1 Aims

1. To test the validity of the brightness continuum to determine whether participants perceive a continuum of test sounds that vary in a monotonic manner in brightness, as is expected if perception is related to the physical differences between the stimuli.

2. To check that sensitivity to differences between sounds from the brightness continuum and the standard varies between low and high sensitivity so that compensation can be measured at low and high sensitivity points.

3. To confirm that the range of the continuum is sufficient to result in high sensitivity performance in future tests conducted with the same continuum under the same conditions.

4. To confirm that the range of the continuum is sufficient to result in high sensitivity performance in future tests with other programme items and other conditions that will be tested in future experiments.

5. To determine whether there is any shift in the perception of the continuum due
to task factors other than precursors (i.e. whether there is any shift due to bias). Shifts caused by factors other than the precursor represent unexpected perceptions and may affect the extent to which compensation can be measured in future tests. These should be removed if possible.

6.4.2 Material

Experiments 4 and 5 (and any future tests) involve testing compensation with music and noise, therefore Pilot 1 tests a brightness continuum that is composed of music. Pilot 2 will run an equivalent test using noise coloured with the LTAS of the music item. In the current pilot test (Pilot 1) a brightness continuum was composed using the Jennifer Warnes Bird on a Wire programme item. This item was chosen for further testing because it is desired to test the same musical programme item as used in Olive et al.’s (1995) real-world experiments to provide a more direct link to that study and directly contribute to explaining compensation in that study. This item is also thought to be conducive to hearing the spectral effects of filtering because of its broad, dense spectrum. All stimuli used in the current experiment consisted of an excerpt of this pop track, taken from the first verse. The excerpt was converted to mono and truncated to 5.1 seconds. This duration was chosen as it encompasses a musical phrase and is long enough to bring about both enhancement and spectral compensation effects, when the excerpt is used to compose precursors in Experiments 4 and 5. In the current test, however, cross-fading was used to create a continuously looping excerpt.

6.4.3 Stimulus Production

A standard sound was produced to act as an unchanging reference throughout the test against which listeners compared comparison sounds for brightness. To make the standard sound the excerpt was filtered with a linear phase high frequency shelf filter with a 1 kHz cut-off in Logic Pro 9. Gain was boosted above 1 kHz by 7 dB to form a midpoint for a continuum of sounds that ranged in brightness.
Table 6.2: Gain values (dB) for filtering above 1kHz. Top row: absolute difference between standard and comparison sound. Following rows: exact dB gain values use to filter the standard and comparison sounds to make Standard/Comparison pairs or S/C intervals.

<table>
<thead>
<tr>
<th>‘S/C interval’</th>
<th>Absolute gain difference (dB)</th>
<th>0.0</th>
<th>1.0</th>
<th>2.0</th>
<th>3.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boost</td>
<td>A Comparison</td>
<td>7.0</td>
<td>8.0</td>
<td>9.0</td>
<td>10.0</td>
</tr>
<tr>
<td></td>
<td>B Standard</td>
<td>7.0</td>
<td>7.0</td>
<td>7.0</td>
<td>7.0</td>
</tr>
<tr>
<td>Cut</td>
<td>C Comparison</td>
<td>-</td>
<td>6.0</td>
<td>5.0</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>D Standard</td>
<td>-</td>
<td>7.0</td>
<td>7.0</td>
<td>7.0</td>
</tr>
</tbody>
</table>

A number of comparison sounds were created to make the brightness continuum. The comparison sounds were created by altering the excerpt spectrally using the same high frequency shelf filter. To create the ‘dark’ and ‘bright’ continuum endpoints HF above 1kHz was boosted by 4 or 10dB, respectively. Intermediate continuum points were created by altering gain above 1kHz in 1dB steps. This resulted in 7 comparison sounds (see Table 6.2 for the exact gains applied). The comparison continuum ranged ±3dB around the midpoint. A 7dB standard stimulus was also created (this was identical to the continuum midpoint). Each of the 7 comparison sounds was paired with the standard to form S/C pairs. This gave 7 S/C pairs and 4 unique differences between the standard and comparison (or 4 S/C intervals—the absolute difference in gain above 1kHz between S and C in a pair). Pre-pilot listening by the experimenter showed the brightness continuum ranged perceptually from ‘darker’ to ‘brighter’ than the standard and it was expected that 50% brighter responses (chance discrimination accuracy) would occur in response to listeners judging the 0 dB S/C interval and that near 0% brighter and 100% brighter responses (100% accuracy) would be obtained when judging the 3 dB S/C interval.

6.4.4 Subjects

Subjects were selected from the University of Surrey’s, Institute of Sound Recording and from its Tonmeister and music-related undergraduate degree programmes. All subjects had extensive experience in critical evaluation of audio quality and most have undergone a technical ear training course. Eight listeners were tested. Each listener judged each S/C pair twice resulting in 16 ratings for each.
6.4 PILOT 1: MUSIC

6.4.5 Method

During the experimental session listeners were seated at a laptop computer within an ITU-R BS 1116 listening room and presented with a page-based experiment interface created in MaxMSP v 4.6. During the test the mono recordings were played diotically to listeners over Sennheiser HD 600 circumaural, open-back headphones. At the beginning of the test listeners were asked to press the play button on the computer interface to begin a sequence of trials. For each trial a new experimental page was presented. During each trial an S/C pair was randomly selected. The standard excerpt and the comparison excerpt looped simultaneously and listeners could switch between hearing one or the other instantaneously by pressing the buttons on the user interface labelled ‘A’ (the standard) or ‘B’ (the comparison). The stimuli were time-aligned to ensure smooth transitions. Listeners where presented with each of the 7 S/C pairs in random order, and were asked to discriminate the comparison from the standard for each pair. A text message informed listeners that they should compare sound ‘B’ to sound ‘A’ for brightness and select the appropriate response button to show whether sound ‘B’ (the comparison) was ‘brighter’ or ‘darker’ than sound A (the standard). Listeners were instructed to listen for the overall spectral tilt of the sound when judging the stimuli and to not focus on listening out for changes within particular narrow frequency bands. They were told that stimuli might differ by only a small amount, which they might not be able to hear and that they should guess if they did not know the answer. This instruction encouraged listeners to respond at chance for the 0dB S/C interval. After listeners had made their judgement for a trial the next trial began after a 10s pause. Over the course of the test every S/C pair was presented to the listener in random order twice.

6.4.6 Results

The validity of the continuum, sensitivity displayed at various continuum points and bias in response to the continuum are discussed.
6.4 PILOT 1: MUSIC

Validity

The brightness continuum was not validated. Listeners reported that the comparison sounds did not vary in what they perceived to be ‘brightness’. This became apparent from reports by listeners early on in testing. Therefore, data from those listeners was removed from analysis and the question was changed. Subsequently, the eight listeners for whom data is presented were asked to discriminate ‘differences in HF content’ rather than ‘brightness’. The question displayed on the test interface was changed to: ‘Does sound ‘B’ (the comparison) contain ‘More’ or ‘Less’ HF content compared to sound ‘A’ (the standard)’. Listeners were asked to make their response by pressing a button labelled ‘More’ or ‘Less’. All listeners reported that they were able to make this decision and perceived the sounds to vary in HF content. All of the following results describe the results for participants conducting the experiment with this HF continuum. Asking listeners for HF judgements rather than brightness judgements is equally appropriate for measuring shifts in the perceived spectrum of comparison sounds, as both tasks are expected to show timbral shifts and compensation in the same way. In fact, the use of this HF continuum is likely to increase the validity of the results because describing the level of perceived HF content is a more specific task than describing brightness and is more directly related to the physical changes in the stimuli.

The perception of HF content for the music-with-vocals continuum is shown in Figure 6.1. This psychometric function plots the proportion of ‘More’ HF compared to the standard responses (y axis) against the continuum points (i.e. the S/C pairs, the x axis). The proportion of ‘More’ responses with 95% confidence intervals can also be seen in Table 6.3). The expected monotonic trend of fewer ‘More’ responses at the 4 dB continuum endpoint to many ‘More’ responses at the 10 dB endpoint occurred. The HF continuum is therefore validated.

Sensitivity

Figure 6.1 shows that sensitivity varied smoothly across the continuum indicating regions of high and low sensitivity. A nearly full psychometric function was produced, with 6% and 94% ‘More’ responses at continuum endpoints. As expected the proportion
of ‘More’ responses at the midpoint is at 50% (chance performance) demonstrating low sensitivity where the comparison sound does not differ from the standard. This data indicates that filter gains of 4 dB and 10 dB (±3 dB from the continuum midpoint) may be sufficient to produce a continuum of responses that range from chance (low sensitivity) to nearly perfect discrimination performance (high sensitivity) with this programme item in future tests of compensation.

![Figure 6.1: music-with-vocals track. The y axis shows the probability of a ‘More’ HF than the standard response. The x axis shows the dB gain applied to the comparison sound (the S/C interval can be calculated by subtracting this from 7 dB). All points on the continuum are shown from 4-10 dB (the standard is 7 dB). Exact probabilities and confidence intervals are displayed in Table 6.3](image)

JND points quantify the sensitivity of listeners to HF change for this continuum. The CBdB(25%) point is at approximately 6.5 dB. Therefore the ‘darker than standard’ JND is 0.5 dB. The CBdB(75%) point is at approximately 8.75 dB. The ‘brighter than standard’ JND is approximately 1.75 dB (these values have been interpolated by eye due to the low resolution of the measures). There is asymmetry in the JND, which indicates bias (see below), but these values show that the JND for this stimulus lies between approximately 0.5 and 1.75 dB, showing that listeners can discern differences in brightness of approximately 1 dB when listening to this programme item. Confidence
Table 6.3: Probability of a ‘More’ HF than the standard response for the music-with-vocals continuum.

<table>
<thead>
<tr>
<th>music-with-vocals gain &gt;1 kHz</th>
<th>P of a ‘More’ HF response</th>
<th>95% CI (L Bound) (Asy. Wald)</th>
<th>95% CI (U Bound) (Asy. Wald)</th>
<th>95% CI (L Bound) (Adj. Wald)</th>
<th>95% CI (U Bound) (Adj. Wald)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 dB</td>
<td>0.06</td>
<td>-0.06</td>
<td>0.18</td>
<td>-0.01</td>
<td>0.30</td>
</tr>
<tr>
<td>5 dB</td>
<td>0.13</td>
<td>-0.04</td>
<td>0.29</td>
<td>0.02</td>
<td>0.37</td>
</tr>
<tr>
<td>6 dB</td>
<td>0.13</td>
<td>-0.04</td>
<td>0.29</td>
<td>0.02</td>
<td>0.37</td>
</tr>
<tr>
<td>7 dB</td>
<td>0.50</td>
<td>0.26</td>
<td>0.75</td>
<td>0.28</td>
<td>0.72</td>
</tr>
<tr>
<td>8 dB</td>
<td>0.50</td>
<td>0.26</td>
<td>0.75</td>
<td>0.28</td>
<td>0.72</td>
</tr>
<tr>
<td>9 dB</td>
<td>0.81</td>
<td>0.62</td>
<td>1.00</td>
<td>0.56</td>
<td>0.94</td>
</tr>
<tr>
<td>10 dB</td>
<td>0.94</td>
<td>0.82</td>
<td>1.06</td>
<td>0.70</td>
<td>1.01</td>
</tr>
</tbody>
</table>

Intervals at the 25% and 75% points are wide so the true population values are not known. However, the consistency of the trend gives support to these JND locations and high sensitivity when judging sounds from this continuum.

To make comparisons of compensation across different tests it is desirable to compare shifts at the same baseline sensitivity points. Therefore, the HF continua presented in all future tests should ensure that specific sensitivity points occur across tests so that comparisons at these points can be made. It is clear that the low sensitivity (50% correct responses) will be reached when discriminating sounds using any continuum that varies around the standard and includes the standard. We can therefore be confident that all future tests using such a continuum will allow for a measurement at low sensitivity. However, it is not clear that the high accuracy at continuum endpoints seen in the current test will be repeated in future tests with a ±3 dB continuum. Further, when using a continuum of this range with different programme items or test conditions sensitivity at continuum endpoints may be reduced. It should be ensured that the ±3 dB range continuum is sufficient to result in high sensitivity performance across tests if compensation at high sensitivity is to be compared. To ensure that the range of the continuum is sufficient to measure compensation at high sensitivity in all tests with all programme items and test conditions the continuum may need to be extended. Alternatively, the highest sensitivity point that can be reached across tests with the current continuum should be ascertained and it should be determined whether this is sufficient for measuring compensation at high sensitivity.
Extending the continuum range to ensure that high sensitivity is reached across all future tests is possible. However, it may not be desirable to do this because increasing the range will not be likely to bring about maximum performance accuracy if stimulus differences are already large enough to be heard clearly. No matter how easy the discrimination task, 100% accuracy may not be achieved because mistakes are made due to inattention to the task and other factors. Mistakes have a large impact on the achieved accuracy with binomial data and small sample sizes (e.g. 1 mistake in 20 results in a score of 95% correct). Extending the continuum would not prevent mistakes, in fact mistakes may happen more readily because extending the continuum brings about a longer task and listener fatigue (Zacharov 2000). Experiments 4 and 5 are likely to be significantly longer than this pilot due to the added precursor, more inter-trial intervals and an increased number of stimuli being measured. Therefore the risk of lower accuracy due to fatigue is greater. In order to maintain accuracy a balance must be struck between increasing the length of the test and ensuring the range of stimuli presented encompasses stimuli that have differences large enough to result in high and low sensitivity performance and points between. To extend the continuum whilst keeping the test short it is possible to remove stimuli from the middle of the continuum. However, as much resolution as possible should be maintained in the middle of the continuum for measuring JNDs and compensation at a variety of sensitivities in future tests.

The current results show high sensitivity to timbral differences with a continuum that ranges between ±3 dB around the standard. However, random sampling variation means that sensitivity may be lower than that measured in the current test in future tests using the same ±3 dB continuum with the same programme item and under the same test conditions. An assessment of performance with this continuum in experiments 4 and 5 and any future tests is made below to determine the minimum level of sensitivity that might be achieved. Further, sensitivity with a ±3 dB range continuum may also be different with different programme items. The adequacy of this continuum with different programme items/conditions that might be tested as part of experiments 4 and 5 or other future tests is also assessed.

The current pilot test revealed discrimination accuracy of 94% at continuum endpoints with this programme item and a continuum ranging ±3 dB. However, this high accuracy
may have occurred through sampling variation and the true population value at continuum endpoints may be lower than 94%. With only 16 ratings at each continuum point and binomial data it is not possible to obtain high confidence that the values in this test represent the true population values. Confidence intervals at each continuum point are wide (see Table 6.3). The Adjusted Wald 95% confidence intervals at the continuum endpoints range from 0-30%, and from 70-100% (Adjusted Wald 95% CI). To gain a more accurate estimate of the true population value at continuum endpoints it is possible to pool the data from both endpoints -3dB and +3dB. This provides an estimate of 94% correct at continuum endpoints with an Adjusted Wald CI of 79 - 99%. An estimation of the bounds in which the true population value lies with 95% confidence can be given by conducting a non-inferiority test (Walker and Nowacki 2010). According to this test the true population level of sensitivity is not below 82%, with 95% confidence \( \Delta = 69\% \), \( 1 - 2\alpha \) CI \( (N = 32) = 82 - 99\% \) (Adjusted Wald values used). Given a true population sensitivity value of not less than 82% at continuum endpoints with 95% confidence it can be stated that this continuum range elicits genuinely high accuracy. However, in future tests with the same continuum under the same conditions, the obtained result may vary from the true population value, due to sampling variation. Given a true population value of 82% it can be shown that values for a repeated test with the same programme item under the same conditions will not fall below 70% with 95% confidence, providing at least 35 participants are measured \( \Delta = 69\% \), \( 1 - 2\alpha \) CI \( (N = 35) = 70 - 91\% \)

This test sound continuum will not produce accuracy at continuum endpoints of lower than 70% in future tests under the same conditions. The maximum performance in discrimination tests with speech under good conditions is frequently 80-90%, (see Section 5.1), where error rates of 4.0-42.6% in phoneme identification tasks in good listening conditions are reported. A musical discrimination task might be expected to be more difficult than a speech discrimination task. 70% may be considered reasonably high accuracy for a musical discrimination task. Further, this lower bound on endpoint accuracy is an extremely conservative estimate of future performance, which is based on the true population value being at the 5% lower limit and the sample value being at the 5% lower limit of the confidence interval of this. It is highly likely that higher accuracy will be achieved across tests in practice. Where higher accuracy than 70% is obtained
across tests this can be used as the high sensitivity comparison point. Therefore, this continuum may be considered appropriate to ensure high accuracy performance in future tests under these conditions. A minimum of 70% accuracy can be expected in future tests with this programme item with a $\pm 3$ dB continuum under the same conditions with at least 95% confidence and higher accuracy is likely.

**Sensitivity with other programme items**

The above assessment shows the minimum value expected with a continuum range $\pm 3$ dB in future tests with the same programme item under the same conditions. However, in future tests the same conditions will not occur. Future tests will involve measuring compensation with different programme items and test conditions to those tested here. With different programme items/conditions a $\pm 3$ dB continuum may not be sufficient to result in at least 70% accuracy across all tests. This issue cannot be addressed by Pilot 1 alone. It is necessary to run pilots for these other programme items/conditions. However, based on the current results some assessment of expected sensitivity in future experiments can be made.

Compared to the other programme items that may be tested (e.g. speech and noise), the music-with-vocals programme item is expected to be the programme item for which it is most difficult to discriminate timbral differences in LTAS. Firstly this item consists of music rather than speech. Speech may benefit from listener familiarity and special mechanisms of perception that make timbral differences easier to discern. Secondly, it is likely to be more difficult to discriminate spectral changes with music or speech compared to noise because it is commonly known that differences in LTAS are easy to determine with noise (perhaps because there is less other distracting information when listening to a static stimulus). Therefore, higher sensitivity is expected with other programme items than that measured here using 3 dB continuum endpoints. Therefore, 70% accuracy remains a conservative estimate of the sensitivity that may be achieved with this continuum across tests with different programme items.

As well as changes in programme item, sensitivity may be reduced by the act of inserting a precursor, regardless of its colouration, because the insertion of a sound between
the sounds being compared is likely to reduce sensitivity due to memory interference (Cowan 1984). Additionally, future tests of the time course of compensation effects may involve extending the time between the standard and the comparison sound, which would lower sensitivity via memory decay (Cowan 1984). Ideally separate pilots to determine sensitivity with a $\pm 3\,\text{dB}$ continuum under these conditions would be conducted. However, a timbral memory test with the same music-with-vocals programme item shows that sensitivity does not reduce with as much as 30 s between precursor and comparison so it is expected that, at least, extending the time between the standard and comparison by up to 30 s will not reduce sensitivity (Pike et al. 2014). Therefore, the 70% point still represents a conservative estimate of sensitivity at continuum endpoints with time separating the standard and comparison sounds.

A specific pilot has not been conducted to examine the effect of memory interference on sensitivity caused by inserting the precursor. This is because it is not possible to obtain this measure. Inserting a spectrally neutral (unfiltered) sound would involve simply a repetition of the standard, and inserting a filtered precursor will cause shifts and represents the same conditions as those that will be seen in the full experiments measuring compensation. Instead, the effect of inserting a neutral precursor on sensitivity is measured in Experiment 4 by examining the effect of the precursor on sensitivity across both high and low frequency precursor conditions (see Section 6.6.10). This experiment shows that the sensitivity is not affected by the act of inserting a precursor in this case (though bias is affected in some cases).

**Monaural test**

Another condition that may reduce sensitivity is monaural listening. Crossaural listening tests are conducted in Experiment 5, during which the listener must compare a standard to a comparison sound monaurally while a precursor is presented to the contralateral ear (see Section 6.7). The act of listening monaurally is expected to lower sensitivity as monaural listening involves less auditory information, so discrimination may be impaired. A pilot was conducted to measure the effect of monaural listening with the music-with-vocals programme item to determine sensitivity at all points on the $\pm 3\,\text{dB}$ continuum. It was expected that sensitivity would reduce with monaural
listening.

The same method as was described in Section 6.4.5 was used \((n = 8)\). However, the test was conducted with the continuum range used in Pilot 2; this encompassed more continuum points within the \(\pm 3\text{dB}\) range and varied around an unfiltered neutral midpoint (see Section 6.5.3 for more details on this continuum). Monaural listening resulted in high sensitivity (100%) at continuum endpoints \(\pm 3\text{dB}\). Due to low participant numbers confidence intervals are wide (see Table 6.4). However, to gain a better estimate of likelihood that the continuum reaches 100% accuracy, as with the binaural test, data for positive and negative continuum endpoints that show 100% accuracy can be pooled to give the same power as the binaural test. This gives a measure of 100% accuracy with a 95% confidence interval of 87-102% (Adjusted Wald). A non-inferiority test shows that the true population value is not below 91% with at least 95% confidence \((\Delta = 69\% , 1 - 2\alpha \text{ CI } (N = 32) = 91 - 102\% )\).

The non-inferiority test shows that it is likely that the true population value responding at continuum endpoints is highly accurate (at least 91%) with a \(\pm 3\text{dB}\) range continuum. Accuracy at continuum endpoints in this test is increased compared to binaural listening and may show genuinely higher accuracy responses with monaural listing. This would be somewhat surprising as monaural listening is expected to be a more difficult task. This may be an indication that with binaural listening the true population value is higher than the lower limit shown in the previous test. However, there may be reasons why listeners perform more accurately with monaural listening. For example the listener may focus more on a monaural task as they are aware that this may be more difficult. High sensitivity measured at other points of the continuum, including the small JNDs, supports the conclusion that the high sensitivity at continuum endpoints is not due to sampling variation.

Although the true population value in the current test is not less that 91% with 95% confidence, the value obtained with this continuum in the same situations in future tests may vary due to sampling variation. Given that the true value is not below 91%, a non-inferiority test can be used to show that sampled values will not be below 70% when at least 35 participants are tested in future work with this continuum under the same conditions \((\Delta = 69\% , 1 - 2\alpha \text{ CI } (N = 35) = 80 - 97\% )\). This shows that
values are not expected to be below 80% in future tests with the same programme item under the same conditions. However, as with the binaural test, repetitions of the exact conditions here are unlikely. Future tests will involve listening with different programme items and with conditions that may be more detrimental to high accuracy, such as listening with a precursor inserted between the stimuli and/or gap between the stimuli. It is not expected that these stimuli will result in lower than 70% accuracy measured under the current conditions, for the same reasons given for the binaural test (music-with-vocals is expected to be the most difficult to discriminate item and accuracy will not drop with separation in time of the standard and comparisons below 30 s). Reduction in accuracy due to the insertion of a precursor cannot be tested here but is tested in Experiment 5. Therefore, this test represents a conservative estimate of future accuracy with monaural listening at continuum endpoints with other programme items and other test conditions.

Figure 6.2: music-with-vocals track—monaural listening. The y axis shows the probability of a ‘More’ HF than the standard response. The x axis shows the dB gain applied to the comparison sound (the S/C interval can be calculated by subtracting this from 0 dB). The continuum range is the same as used in Pilot 2. Exact probabilities and confidence intervals are displayed in Table 6.4.
Table 6.4: Probability of a ‘More’ response for the music-with-vocals continuum with monaural comparisons.

<table>
<thead>
<tr>
<th>Music: gain &gt;1 kHz</th>
<th>P of a ‘More’ HF response</th>
<th>95% CI (L Bound)</th>
<th>95% CI (U Bound)</th>
<th>95% CI (Asy. Wald)</th>
<th>95% CI (Adj. Wald)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3.0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.05</td>
<td>0.37</td>
</tr>
<tr>
<td>-1.8</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.05</td>
<td>0.37</td>
</tr>
<tr>
<td>-1.6</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.05</td>
<td>0.37</td>
</tr>
<tr>
<td>-1.4</td>
<td>0.25</td>
<td>-0.05</td>
<td>0.55</td>
<td>0.06</td>
<td>0.60</td>
</tr>
<tr>
<td>-1.2</td>
<td>0.00</td>
<td>-0.05</td>
<td>0.55</td>
<td>-0.05</td>
<td>0.37</td>
</tr>
<tr>
<td>-1.0</td>
<td>0.25</td>
<td>-0.05</td>
<td>0.55</td>
<td>0.06</td>
<td>0.60</td>
</tr>
<tr>
<td>-0.8</td>
<td>0.25</td>
<td>-0.05</td>
<td>0.55</td>
<td>0.06</td>
<td>0.60</td>
</tr>
<tr>
<td>-0.6</td>
<td>0.25</td>
<td>-0.05</td>
<td>0.55</td>
<td>0.06</td>
<td>0.60</td>
</tr>
<tr>
<td>-0.4</td>
<td>0.50</td>
<td>0.15</td>
<td>0.85</td>
<td>0.22</td>
<td>0.79</td>
</tr>
<tr>
<td>-0.2</td>
<td>0.50</td>
<td>0.15</td>
<td>0.85</td>
<td>0.22</td>
<td>0.79</td>
</tr>
<tr>
<td>0.0</td>
<td>0.38</td>
<td>0.13</td>
<td>0.61</td>
<td>0.18</td>
<td>0.61</td>
</tr>
<tr>
<td>0.2</td>
<td>0.75</td>
<td>0.45</td>
<td>1.05</td>
<td>0.40</td>
<td>0.94</td>
</tr>
<tr>
<td>0.4</td>
<td>0.75</td>
<td>0.45</td>
<td>1.05</td>
<td>0.40</td>
<td>0.94</td>
</tr>
<tr>
<td>0.6</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.63</td>
<td>1.05</td>
</tr>
<tr>
<td>0.8</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.63</td>
<td>1.05</td>
</tr>
<tr>
<td>1.0</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.63</td>
<td>1.05</td>
</tr>
<tr>
<td>1.2</td>
<td>0.75</td>
<td>0.45</td>
<td>1.05</td>
<td>0.40</td>
<td>0.94</td>
</tr>
<tr>
<td>1.4</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.63</td>
<td>1.05</td>
</tr>
<tr>
<td>1.6</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.63</td>
<td>1.05</td>
</tr>
<tr>
<td>1.8</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.63</td>
<td>1.05</td>
</tr>
<tr>
<td>3.0</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.63</td>
<td>1.05</td>
</tr>
</tbody>
</table>
Bias

Another aim of this experiment was to confirm that there are no CB shifts away from expected values that indicate bias in the perception of the continuum. Only the binaural data is examined in this analysis. The continuum varies in HF content around 7 dB in a symmetrical manner so it is expected that responses will also vary around 7 dB in a symmetrical manner. Bias can be determined by the shape and position of the psychometric function along the x axis and by CB measures. Slight bias can be seen by examining Figure 6.1. It is evident that there is a shift in the continuum towards the right hand side of the x axis indicating a tendency to report fewer ‘More’ HF responses than expected. Table 6.5 describes the shift in the continuum using all CB measures. Almost all measures showed bias toward fewer ‘More’ HF than the standard responses. The CB% overall is 37%. This is the most powerful measure of bias as all data points are included in this measure. This value is significantly different from the expected 50% ‘More’ responses ($Z = 2.0, p = .0479$, two tailed). The CBdB(50%) is at 7.5 dB and this also indicates slight bias towards fewer ‘More’ HF responses than expected (the expected value is 7 dB), significance cannot be measured for CBdB points. The CBdB(25%) is at approximately 6.5 dB (the ‘darker than standard’ JND is therefore 0.5 dB) and the CBdB(75%) point is at approximately 8.75 dB (the ‘brighter than standard’ JND is approximately 1.75 dB). The asymmetry is indicative of increased sensitivity on the dark side of the continuum and bias towards fewer ‘More’ HF responses than expected. However, no bias is shown by the CB%(midpoint) measure which is 50%, as expected. The true bias at any single point is not estimated with certainty, as binomial data and small sample sizes result in large confidence intervals, but the consistency of bias across the continuum supports a conclusion of bias, as does the significance of the overall measure which is highly powered. Bias towards hearing the comparison sounds as darker than the standard is not expected and methods of removing this baseline bias should be considered, as this may effect the measurement of shifts caused by compensation.
Table 6.5: Pilot 1: Bias measures for the music-with-vocals continuum

<table>
<thead>
<tr>
<th>Measure</th>
<th>N</th>
<th>Expected</th>
<th>Result</th>
<th>Bias?</th>
<th>Significance of difference (2 tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CB% (overall)</td>
<td>Proportion of ‘More’ responses across the whole continuum.</td>
<td>50%</td>
<td>37%</td>
<td>Bias towards fewer ‘More’ responses</td>
<td>$Z = 2, \ p = 0.0497$</td>
</tr>
<tr>
<td>CB% (midpoint - 7 dB)</td>
<td>Proportion of ‘More’ at 7 dB comparison sound.</td>
<td>50%</td>
<td>50%</td>
<td>No bias</td>
<td>-</td>
</tr>
<tr>
<td>CBdB(50%)</td>
<td>The location of 50% ‘More’ HF responding</td>
<td>7 dB</td>
<td>7.5 dB</td>
<td>Bias towards fewer ‘More’ responses</td>
<td>-</td>
</tr>
<tr>
<td>CBdB(25%)</td>
<td>The location of 25% ‘More’ HF responding</td>
<td>Symmetrical with CBdB(75%)</td>
<td>6.5 dB (JND = 0.5 dB)</td>
<td>Bias towards fewer ‘More’ responses</td>
<td>-</td>
</tr>
<tr>
<td>CBdB(75%)</td>
<td>The location of 75% ‘More’ HF responding</td>
<td>Symmetrical with CBdB(25%)</td>
<td>8.75 dB (JND = 1.75 dB)</td>
<td>Bias towards fewer ‘More’ responses</td>
<td>-</td>
</tr>
</tbody>
</table>
6.4 PILOT 1: MUSIC

6.4.7 Summary and discussion

A continuum ranging in brightness was not validated but stimuli were heard to vary monotonically in HF content, from ‘Less’ to ‘More’ HF compared to the standard. The task was therefore changed to one in which listeners discriminated the HF content of comparison sounds compared to a standard sound. All future tests of compensation will involve discriminating HF content rather than brightness. The expected trend of increased ‘More’ HF responses with increased gain above 1 kHz and decreased ‘More’ responses with decreased gain above 1 kHz was confirmed and the HF continuum was validated. As expected at the continuum endpoints there was high sensitivity to HF differences and listeners responded with near perfect discrimination accuracy at these points (94%). As expected low sensitivity (50% responses) was seen at the continuum midpoint. The JND for changes in HF above 1 kHz with this programme item, under these listening conditions, is approximately 1 dB.

It is acknowledged that compensation may be larger at some sensitivities than others and this continuum may be sufficient to allow for an assessment of compensation at the same high and low sensitivity points across tests. However, the current test alone was not sufficient to show that a full range of sensitivity will be produced with a continuum ranging ±3 dB in future tests as the current results may be a result of sampling variation. A non-inferiority test shows that the true population value for accuracy at continuum endpoints may be as low as 82% with 95% confidence; this is still considered high accuracy. However, given that the true population value is 82%, future tests using this continuum may produce lower values than this due to sampling variation. A non-inferiority test showed that values are highly unlikely to be lower than 70% with any given sample providing a sample size of at least 35. 70% is reasonably high accuracy considering that tests with vowel stimuli that are thought to be easier to discriminate than non-speech stimuli, frequently result in around 80% accuracy under good conditions. If at least 70% accuracy can be achieved at continuum endpoints across all tests then this is sufficient to allow for a comparison of compensation at both low and high sensitivity points. Ideally a comparison of compensation across tests at least 80% accuracy is desirable for comparing compensation at high sensitivity. Because of the non-inferiority tests resulting in in highly conservative estimates, it is
likely that in practice values of 80% or higher will be achieved at continuum endpoints across tests with this continuum and this programme item under the same conditions.

In tests involving different programme items and conditions sensitivity at $\pm 3\text{dB}$ continuum endpoints may reduce further. However, it was noted that it is not likely to reduce further with different programme items as the music-with-vocals represents the most difficult to discriminate item that is likely to be tested. In fact too high accuracy may be a problem with other programme items (e.g. noise, see Section 6.5).

It was shown that crossaural precursor tests conducted as part of Experiment 5 will not result in reduce sensitivity due to the monaural nature of the comparisons made. It was not shown that the effect of inserting a precursor between the standard and comparison will not reduce sensitivity due to memory interference (but this is shown in Experiment 4—see Section 6.6.10). It was shown that the act of separating the standard and comparison by up to 30 s will not reduce sensitivity due to memory decay.

Therefore, this continuum appears to contain an appropriate range of S/C intervals to yield high (at least 70%) and low sensitivity in future tests with the same programme item under the sort of conditions that may be tested in future work.

Bias in the continuum may be detrimental to accurately observing compensation as bias causes a shift in the continuum in the same manner as compensation and may prevent further shifts caused by compensation due to ceiling effects. Any shifts of the continuum away from the midpoint on the x axis in this pilot experiment represents unexpected bias caused by extraneous factors. The overall measure (which has the greatest statistical power) showed significant bias toward fewer ‘More’ responses than expected and other measures revealed consistent evidence of this bias. Attempts should be made to discover the source of this bias and remove it. It is not clear why a bias occurred. It could be caused by the perceived naturalness of comparison sounds. Listeners might perceive the spectrum of the original track to be a natural spectrum for this sort of pop music (particularly if they have heard the track before). All the stimuli in this test were boosted above 1 kHz compared to the original programme item. This shift towards making the timbre of the track brighter than the original programme item might cause asymmetry in the perception of sounds. Future studies should consider using a natural, unfiltered, midpoint that contains the original LTAS of the item; this may remove the bias. Ideally programme items that the listeners have
not heard before should be used in future testing to avoid perceptions from outside the experiment from influencing the task. An alternative explanation of the bias is that listeners were responding ‘Less’ when they did not know the correct answer, but no listeners indicated that this occurred.

6.4.8 Pilot 1 Conclusion

The aim of this test was to determine whether the brightness continuum is suitable for measuring compensation in Experiments 4 and 5 and future work. How the aims of this test were fulfilled is summarised.

Aim 1: To validate the brightness continuum

Listeners did not perceive the continuum as varying in brightness. Therefore, tests were conducted asking participants to determine whether the same comparison sounds contained ‘More’ or ‘Less’ high frequency content compared to the standard. Listeners perceived the sounds to vary in HF and psychometric functions showed that the stimuli varied in perceived HF content in the monotonic manner expected by physical changes to the stimuli. Therefore a HF continuum, that is appropriate for testing compensation, was validated.

Aim 2: To check that discrimination sensitivity varies across the continuum between low and high sensitivity so that compensation can be measured at different sensitivity points.

The extent to which compensation occurs may vary depending on baseline sensitivity, so compensation should be measured at a number of baseline sensitivities. Ideally, all tests would measure compensation at 50% (low) and 100% (perfect) accuracy and points between this. All tests that involve the comparison of sounds from a continuum that varies around a midpoint sound which is identical to the standard will produce low sensitivity somewhere along the continuum (where there is no bias this is expected to occur at the point on the continuum that is the same as the standard sound). Therefore, all tests described in this chapter will be able to measure compensation at low sensitivity and it is
expected that this will be at or near the continuum midpoint. It was seen that 50% sensitivity was achieved at the continuum midpoint (at 7 dB) in this test, as expected. However, it is acknowledged that 100% (perfect) accuracy is not expected to occur at continuum endpoints in all tests. Even high accuracy may be difficult to achieve across a range of conditions. Extending the continuum will not necessarily ensure high accuracy. A lower point than perfect accuracy may be considered high enough accuracy for the measurement of compensation at high sensitivity. The current continuum revealed that true accuracy at continuum endpoints was not below 82%. This is high enough to be considered high accuracy.

Aim 3: To confirm that the range of the continuum is sufficient to result in high sensitivity performance in future tests conducted with the same continuum under similar conditions

High accuracy is expected at continuum endpoints. The true accuracy at continuum endpoints is high (82%). However, when tests are run with the current continuum in the future under the same conditions measured values may vary across tests due to sampling variation. Given a true population accuracy of 82%, values of lower than 70% are not expected across tests due to sampling variation. This is deemed high enough to be suitable for measuring compensation at high accuracy. However, in practice higher accuracy is very likely to be reached at continuum endpoints. If higher than 70% accuracy is reached across all tests then this higher accuracy point may be used for comparison.

Aim 4: To confirm that the range of the continuum is sufficient to result in high sensitivity performance in other conditions to be tested as part of a series of experiments.

High accuracy is expected at continuum endpoints. Variations in programme item and conditions in future tests mean that a ±3 dB continuum accuracy may result in accuracy below the 70% minimum estimated here. An assessment of other programme items and conditions that will be measured in Experiments 4 and 5 and future experiments was made. This assessment revealed that sensitivity is unlikely to fall below 70%. In particular, changing the programme item, moving
6.5 PILOT 2: NOISE

the stimuli apart in time, inserting precursors and monaural listening that may occur in future tests is not likely to reduce sensitivity at the continuum endpoints to below 70%. It can be concluded that a continuum that ranges ±3 dB around the midpoint is suitable for future testing with a range of programme items and conditions where high (at least 70%) and low (chance) sensitivity performance is desired.

Aim 5: **To determine whether there is any shift in the perception of the continuum, due to task factors other than precursors**

The continuum varies symmetrically around the midpoint so bias is not expected. However, listeners displayed a significant tendency to rate stimuli as containing less HF content compared to the standard. This bias is consistent and worthy of further investigation as it may affect the measurement of compensation in future tests via ceiling effects. This bias is suggestive of something irregular about the perception of the continuum which should be addressed. The bias may be a result of using a spectrally *unnatural* mid-point for the continuum. A natural midpoint, where the timbre of the comparison sounds vary around an unfiltered standard should be used in future tests to attempt to eliminate this bias.

### 6.5 Pilot 2: Noise

A further pilot test was carried out with the noise programme item as it was expected that sensitivity might be higher with this programme item compared to music; this may make it difficult to obtain measures of compensation at middle sensitivity points and measures of JNDs.

#### 6.5.1 Aims

The aims of this pilot are the same as for Pilot 1: to test the validity, sensitivity and bias of the HF continuum when presented without precursors (see Section 6.4.1). However, emphasis is placed on confirming that a full psychometric function is observed with an
adequate range of sensitivities at middle points on the continuum.

6.5.2 Material

A noise programme item that has the same LTAS as the music item (Jennifer Warnes: Bird on a Wire) was composed by producing white noise using the test oscillator plugin in Logic Pro 9 and applying the LTAS of the music excerpt using Match Eq Logic Pro 9. This plugin calculates the LTAS of the music segment over the course of the whole excerpt and applies this, via filtering, to the noise sample.

6.5.3 Stimulus Production

The standard and the comparison sounds were produced in the same manner as described in Section 6.4.3, but filtering values were different. A standard sound was created that acted as an unchanging reference throughout the test against which listeners compared comparison sounds. Unlike in Pilot 1 this standard was not filtered above 1kHz, so the standard contained the natural LTAS of the music item and represented a natural continuum midpoint. This was done to reduce the possibility of the bias mentioned at the end of Section 6.4.7.

The comparison sound continuum consisted of stimuli that varied ±3 dB in HF content above 1 kHz around this 0 dB gain natural midpoint. Pre-pilot listening showed increased sensitivity to this programme item compared to the music programme item and it was apparent that ±3 dB endpoints would be likely to result in 100% accuracy. In light of the potential increase in sensitivity the continuum was made to encompass more points within the ±3 dB range. Comparison sounds were composed that varied in gain above 1 kHz between ±3 dB in 0.2 dB steps up to 2 dB and thereafter 0.5 dB (see Table 6.6 for the exact gains applied). 2.0 dB and 2.5 dB points were removed from the continuum to create a shorter test. The 0 dB comparison stimuli formed the continuum midpoint, which was the same as the standard. Each of the comparison sounds was paired with the standard to form 22 S/C pairs (11 S/C intervals). These can be seen in Table 6.6. The 0 dB point was measured twice to gain more data at this point.
Pre-pilot listening by the experimenter showed this continuum ranged perceptually from ‘Less’ to ‘More’ HF compared to the standard.
Table 6.6: Top row: absolute difference between standard and comparison sound. Following rows: exact dB gain values used to filter the comparison and standard sounds to make S/C pairs or S/C intervals.

<table>
<thead>
<tr>
<th>S/C interval difference (dB)</th>
<th>Boost A</th>
<th>B Comparison</th>
<th>Standard C</th>
<th>D Comparison</th>
<th>Standard D</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>0.2</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>0.4</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>0.6</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.8</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>1.2</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>1.4</td>
<td>1.4</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>1.6</td>
<td>1.6</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>1.8</td>
<td>1.8</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>3.0</td>
<td>3.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
</tbody>
</table>
6.5.4 Subjects

The selection criteria were the same as described in Section 6.4.4. Seven listeners were tested.

6.5.5 Method

The method is the same as that described for Pilot 1 (see Section 6.4.5). Listeners were asked to judge whether the comparison stimuli taken from all continuum points contained ‘More’ or ‘Less’ HF content compared to the standard by pressing buttons labelled ‘More’ or ‘Less’ on a computer interface. Listeners were presented with each S/C pair in a random order.

6.5.6 Results

The validity of the continuum, sensitivity displayed at various continuum points and bias in response to the continuum is discussed.

Validity

The HF continuum was validated. Listeners reported that the sounds from the comparison continuum did vary around the standard in what they perceived to be HF content. Participants reported that the stimuli sounded like pink noise. This is expected based on the spectral centroid of this stimulus. The perception of HF content for the noise continuum is shown in Figure 6.3. This psychometric function plots the proportion of ‘More HF compared to the standard’ responses or ‘More’ responses (y axis) against the continuum points (i.e. the S/C pair, the x axis). The proportion of ‘More’ responses with 95% confidence intervals can also be seen in Table 6.7. The expected monotonic trend of fewer ‘More’ responses at the -3 dB continuum endpoint to many ‘More’ responses at the 3 dB endpoint occurred. The continuum was validated.
### Table 6.7: Probability of a ‘More’ HF response for the Noise-with-the-LTAS-of-music item continuum.

<table>
<thead>
<tr>
<th>Noise: gain &gt;1kHz</th>
<th>P of a ‘More HF’ response</th>
<th>95% CI (L Bound)</th>
<th>95% CI (U Bound)</th>
<th>95% CI (L Bound)</th>
<th>95% CI (U Bound)</th>
<th>95% CI (L Bound)</th>
<th>95% CI (U Bound)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3.0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.05</td>
<td>0.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.8</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.05</td>
<td>0.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.6</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.05</td>
<td>0.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.4</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.05</td>
<td>0.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.2</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.05</td>
<td>0.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.05</td>
<td>0.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.8</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.05</td>
<td>0.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.6</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.05</td>
<td>0.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.4</td>
<td>0.14</td>
<td>-0.12</td>
<td>0.40</td>
<td>0.05</td>
<td>0.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.2</td>
<td>0.29</td>
<td>-0.05</td>
<td>0.62</td>
<td>0.08</td>
<td>0.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>0.36</td>
<td>0.11</td>
<td>0.61</td>
<td>0.16</td>
<td>0.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>0.57</td>
<td>0.21</td>
<td>0.94</td>
<td>0.25</td>
<td>0.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>0.71</td>
<td>0.38</td>
<td>1.05</td>
<td>0.35</td>
<td>0.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.60</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.60</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.86</td>
<td>0.60</td>
<td>1.12</td>
<td>0.47</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.60</td>
<td>1.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.60</td>
<td>1.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.6</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.60</td>
<td>1.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.8</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.60</td>
<td>1.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.0</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.60</td>
<td>1.05</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 6.3: Noise track. The y axis shows the probability of a ‘More’ HF than the standard response. The x axis shows the dB gain applied to the comparison sound (the S/C interval can be calculated by subtracting this from 0 dB). All points on the continuum are shown from -3 to +3 dB (the standard is 0 dB). Exact probabilities and confidence intervals are displayed in Table 6.3

**Sensitivity**

Figure 6.3 shows that a full psychometric function was produced by the HF continuum, with 0 and 100% ‘More’ responses at continuum endpoints. The psychometric function is steeper for this programme item compared to the music-with-vocals programme item, showing that listeners are more sensitive to differences in HF content above 1 kHz when listening to noise compared to music; the psychometric function is also smoother. The JND measures show increased discrimination accuracy of HF content. The CBdB(25%) point is approximately -0.2 dB (the ‘Less HF than the standard’ JND is 0.2 dB). The CBdB(75%) point is approximately 0.4 dB (the ‘More HF than the standard’ JND is 0.4 dB). Overall the JND is approximately 0.3 dB. Sensitivity is therefore increased by approximately 0.7 dB compared with music. It is not possible to be highly confident that JNDs represent true sensitivity because confidence intervals at these specific points are large (see Table 6.7). However, the trend of high sensitivity across the continuum supports the conclusion that this continuum results in a small JND. It can also be observed that there is sufficient range of accuracy between the 0 dB points and endpoints to measure compensation at low and high sensitivity and
points with this programme item. However, it may be preferable to include more points within the continuum to ensure the measurement of compensation at a wider range of sensitivities in future tests.

This data indicates that a continuum that ranges ±3 dB from the continuum midpoint is sufficient to produce responses that range from 50% accuracy (chance) to 100% accurate responses with this programme item under these conditions. However, based on this test alone it cannot be concluded that 100% accuracy will be reached in future tests with this programme item under these conditions due to sampling variation and conditions which may make accurate performance more difficult. To gain a more accurate estimate of the true population value at continuum endpoints it possible to pool the data from the positive and negative endpoints in the same way as is done for the music-with-vocals item. This provides an estimate of 100% correct at continuum endpoints with an Adjusted Wald 95% CI's of 86-102%

An assessment of the minimum values expected to be achieved at continuum endpoints in future tests with the same programme item under the same conditions with the same ±3dB continuum is made using non-inferiority testing (Walker and Nowacki 2010). Given a sample accuracy of 100% at continuum endpoints, this test reveals that true population responding is not lower than 90% with 95% confidence (\(Delta = 69\), \(1 - 2\alpha CI \ (N = 28) = 90 - 102\%\)). This shows very high accuracy is obtained with a ±3dB continuum. However, future tests using this continuum with this programme item under the same conditions may yield a value lower than this due to random sampling variation. Given a true value of not less than 90%, responding is not expected to be lower than 77% with 95% confidence providing at least 35 participants are tested (\(Delta = 69\), \(1 - 2\alpha CI \ (N = 35) = 77 - 95\%\)). Therefore, it is highly likely that the at least 70% accuracy will be achieved at continuum endpoints with a ±3dB continuum made up of this programme item under these conditions.

Whilst performance of at least 70% at continuum endpoints is confirmed with high confidence for the current programme item and current conditions, future tests are not likely to be conducted under the same conditions. Accuracy may vary when different programme items and conditions are used, for the same reason as is discussed with the music programme item (see Section 6.4.6). Noise is expected to be the programme
item that is most conducive to hearing differences in LTAS. As was discussed in Pilot 1 high sensitivity at continuum endpoints is less likely for other programme items but that sensitivity and at least 70% accuracy is likely to be obtained for those programme items. Therefore, it is even more likely that the 70% point or higher will be reached with the noise programme item. Further, in future test stimuli may need to be separated in time. Like for the music item the study by Pike et al. (2014) shows that sensitivity does not decrease further with separation of the standard from the comparison sound up to 30 s with the same noise programme item as tested here. As was discussed for the music programme item, it is not possible to measure the effect of inserting a precursor on sensitivity with a pilot test, but this is tested in Experiment 4. It is possible that sensitivity may be reduced or increased with tests involving monaural presentation (as will be run as part of Experiment 5). A test with monaural listening was therefore conducted to examine this.

**Monaural test**

A monaural test was conducted to determine whether monaural listening to the noise continuum lowers sensitivity. Selection criteria were the same as for the music pilot described in Section 6.4.4 and 7 participants were tested. Results are shown in Figure 6.4. It appears that this condition reveals that high sensitivity is reached within ±3 dB about the standard. As for the music item the measured accuracy is slightly higher with monaural listening compared to binaural listening. There is some asymmetry in JNDs but the (CBdB(25%) = 0.3 dB (‘Less HF than standard’ JND is 0.3 dB). The CBdB(75%) is 0.1 dB (‘More HF than standard’ JND is 0.1 dB). Therefore the JND is approximately 0.2 dB for HF change above 1 kHz. As the JND is slightly smaller than with binaural listening it may be preferable to add more points between 0 and 2 dB to the continuum to obtain better resolution for measuring JNDs and compensation in the middle of the continuum in future experiments.

As with the binaural data, the data from the endpoints can be pooled giving the same confidence intervals as with the binaural data and the same non-inferiority test results (due to the same given accuracy level and participant numbers in this test as the binaural test). Therefore, it can be concluded that at least 70% accuracy will be
reached using this continuum range with this programme item in these conditions in future tests. For the reasons noted in the binaural noise test (see Section 6.5.6), a further reduction in accuracy is not expected with other programme items and test conditions.

![Figure 6.4: Noise track—monaural listening. The y axis shows the probability of a ‘More’ HF than the standard response. The x axis shows the dB gain applied to the comparison sound (the S/C interval can be calculated by subtracting this from 0 dB). Exact probabilities and confidence intervals are displayed in Table 6.8](image-url)
Table 6.8: Probability of a ‘More’ HF response for the Noise continuum with monaural comparisons.

<table>
<thead>
<tr>
<th>Noise: gain &gt;1 kHz</th>
<th>P of a ‘More HF’ response</th>
<th>95% CI (L Bound)</th>
<th>95% CI (U Bound)</th>
<th>95% CI (Asy. Wald)</th>
<th>95% CI (Adj. Wald)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3.0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.05</td>
<td>0.40</td>
</tr>
<tr>
<td>-1.8</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.05</td>
<td>0.40</td>
</tr>
<tr>
<td>-1.6</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.05</td>
<td>0.40</td>
</tr>
<tr>
<td>-1.4</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.05</td>
<td>0.40</td>
</tr>
<tr>
<td>-1.2</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.05</td>
<td>0.40</td>
</tr>
<tr>
<td>-1.0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.05</td>
<td>0.40</td>
</tr>
<tr>
<td>-0.8</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.05</td>
<td>0.40</td>
</tr>
<tr>
<td>-0.6</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.05</td>
<td>0.40</td>
</tr>
<tr>
<td>-0.4</td>
<td>0.14</td>
<td>0.04</td>
<td>0.58</td>
<td>0.05</td>
<td>0.53</td>
</tr>
<tr>
<td>-0.2</td>
<td>0.29</td>
<td>-0.05</td>
<td>0.62</td>
<td>0.08</td>
<td>0.65</td>
</tr>
<tr>
<td>0.0</td>
<td>0.64</td>
<td>0.39</td>
<td>0.90</td>
<td>0.39</td>
<td>0.84</td>
</tr>
<tr>
<td>0.2</td>
<td>0.86</td>
<td>0.60</td>
<td>1.12</td>
<td>0.47</td>
<td>1.00</td>
</tr>
<tr>
<td>0.4</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.60</td>
<td>1.05</td>
</tr>
<tr>
<td>0.6</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.60</td>
<td>1.05</td>
</tr>
<tr>
<td>0.8</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.60</td>
<td>1.05</td>
</tr>
<tr>
<td>1.0</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.60</td>
<td>1.05</td>
</tr>
<tr>
<td>1.2</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.60</td>
<td>1.05</td>
</tr>
<tr>
<td>1.4</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.60</td>
<td>1.05</td>
</tr>
<tr>
<td>1.6</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.60</td>
<td>1.05</td>
</tr>
<tr>
<td>1.8</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.60</td>
<td>1.05</td>
</tr>
<tr>
<td>3.0</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.60</td>
<td>1.05</td>
</tr>
</tbody>
</table>
Table 6.9: Bias measures for the Noise programme continuum

<table>
<thead>
<tr>
<th>Measure</th>
<th>Expected</th>
<th>Result</th>
<th>Bias?</th>
<th>Significance of difference (2 tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CB% (overall) Proportion of ‘More’ responses across the whole continuum.</td>
<td>50%</td>
<td>47%</td>
<td>Bias towards fewer ‘More’ responses</td>
<td>NS $Z = 0.5, p = 0.598$</td>
</tr>
<tr>
<td>CB% (0 dB) Proportion of ‘More’ responses for the 0 dB comparison sound.</td>
<td>50%</td>
<td>36%</td>
<td>Bias towards fewer ‘More’ responses</td>
<td>$Z = 0.7, p = .454$</td>
</tr>
<tr>
<td>CBdB(50%) The location of 50% ‘More’ HF responding</td>
<td>0 dB</td>
<td>0.2 dB</td>
<td>Bias towards fewer ‘More’ responses</td>
<td>-</td>
</tr>
<tr>
<td>CBdB(25%) The location of 25% ‘More’ HF responding</td>
<td>Symmetrical with CBdB(75%)</td>
<td>-0.2 dB (JND = -0.2 dB)</td>
<td>Bias towards fewer ‘More’ responses</td>
<td>-</td>
</tr>
<tr>
<td>CBdB(75%) The location of 75% ‘More’ HF responding</td>
<td>Symmetrical with CBdB(25%)</td>
<td>0.4 dB (JND = 0.4 dB)</td>
<td>Bias towards fewer ‘More’ responses</td>
<td>-</td>
</tr>
</tbody>
</table>
Bias

Another aim of Pilot 2 was to confirm that there was no CB shift away from expected values (bias), as this may affect the extent to which compensation can occur and represents an unexpected response to the symmetrical continuum. Only the binaural data is examined in this analysis. Bias can be determined by the shape and position of the psychometric function along the x axis and by CB measures. Slight bias can be seen by examining Figure 6.3. There appears to be a small shift in the continuum towards the right-hand side of the x axis, indicating a tendency to report fewer ‘More’ responses than expected.

Table 6.13 describes the shift in the continuum with the range of CB measures set out above. Almost all show slight bias toward fewer ‘More’ HF than the standard responses. The CB%(overall) is 47% which is slightly different from the expected 50% value. This is the most powerful measure of bias as it is based on all data points; however, this value is not significantly different from 50% ‘More’ responses. The CB%(0 dB) measure at 0.2 dB also indicates slight bias towards fewer ‘More’ HF responses than expected but due to low power and small bias this is not statistically significant. The CBdB(25%) is at approximately -0.2 dB (The ‘Less HF than standard’ JND is therefore 0.2 dB) and the CBdB(75%) point is at approximately 0.4 dB (the ‘More HF than standard’ JND is approximately 0.4 dB). The asymmetry is indicative of increased sensitivity on the ‘Less HF than the standard’ side of the continuum and bias towards fewer ‘More’ HF responses.

On balance there is some evidence of bias across measures towards hearing the stimuli to contain less HF than the standard. The reasons for this are not clear. The result may be to do with listeners preferring to respond ‘Less’ for reasons other than the perception of HF content, such as a tendency to report ‘Less’ when they did not know the answer. However, listeners claimed they were not responding in this way. As the overall measure is far from significant and there is no apparent explanation for this it may be concluded that this bias may be a result of random variation and no steps will be taken to further reduce this. It is evident that the bias is not large enough to have an impact on compensation in future tests as it does not cause ceiling effects.
6.5.7 Pilot 2 Conclusion

The aim of this test was to determine whether the noise-based HF continuum is suitable for measuring compensation in Experiments 4 and 5 and future work. How the aims of this test were fulfilled is summarised.

Aim 1: To validate the continuum

Listeners perceived noise coloured by the LTAS of the music programme item, which varied in HF above 1 kHz by $\pm 3\text{dB}$, as varying in HF content around the standard in the monotonic manner expected by physical changes to the stimuli. Full psychometric functions were seen and a HF continuum made up of noise was validated.

Aim 2: To check that discrimination sensitivity varies across the continuum between low and high sensitivity so that compensation can be measured at different sensitivity points.

The extent to which compensation occurs may vary depending on baseline sensitivity, so compensation should be measured at a number of baseline sensitivities. Ideally, all tests would measure compensation at 50% (low) and 100% (perfect) accuracy and points between this. As already observed in Section 6.4.8, all tests that involve the comparison of sounds from a continuum that varies around a midpoint sound which is identical to the standard will produce low sensitivity somewhere along the continuum (where there is no bias this is expected to occur at the point on the continuum that is the same as the standard sound). Therefore, all tests described in this chapter will be able to measure compensation at low sensitivity and it is expected that this will be at or near the continuum midpoint. It was seen that 50% sensitivity was achieved near to the continuum midpoint in this test (at 0.2 dB). However, it is acknowledged that 100% (perfect) accuracy is not expected to occur at continuum endpoints. Even obtaining high accuracy performance may be difficult. Extending the continuum will not necessarily ensure high accuracy. A lower point may be considered high enough accuracy for the measurement of compensation at high sensitivity. The current continuum revealed 100% discrimination accuracy in
the current test with a continuum ranging at least ±3 dB. Non-inferiority tests showed that the true population response value may be as low as 90%. The true value can still be regarded as high accuracy.

**Aim 3:** To confirm that the range of the continuum is sufficient to result in high sensitivity performance in future tests conducted with the same continuum under the same conditions.

High accuracy is expected at continuum endpoints. True accuracy is high (at least 90%) when tests are run with the current continuum under the current conditions but sampling variation may mean lower accuracy responses are obtained in future tests with the same continuum under the same conditions. It was confirmed that at least 70% accuracy would be reached with this continuum under these conditions in future tests.

**Aim 4:** To confirm that the range of the continuum is sufficient to result in high sensitivity performance in other conditions to be tested as part of a series of experiments.

High accuracy is expected at continuum endpoints. Variations in programme items and conditions in future tests mean that a ± 3 dB continuum accuracy may result in reduced accuracy at continuum endpoints compared to that estimated here. With the insertion of the precursor and extending time gaps monaural listening sensitivity may decrease further, but it is concluded that accuracy is not expected to be below 70% with 95% confidence (the effect of inserting precursors was not tested). While the range of sensitivities is appropriate with this continuum, too accurate responding may be a problem. Sensitivity to differences is notably increased with noise stimuli compared to the music. The JND was small (0.3 dB). This means that it may be difficult to measure compensation at the mid sensitivity points. It may be advisable to include more continuum points between ± 2 dB to allow for increased sensitivity around the mid sensitivity region of the continuum.

**Aim 5:** To determine whether there is any shift in the perception of the continuum, due to task factors other than precursors
Bias may cause ceiling effects and is suggestive of something irregular about the perception of the continuum, which should be addressed. The continuum varies symmetrically around the midpoint so bias is not expected. However, listeners displayed a non-significant tendency to rate stimuli as containing less HF content compared to the standard. The test used a natural midpoint and listeners did not report a reason for reporting ‘Less’ more frequently. It is not clear that any experimental factors are causing this bias. As the measure showed this bias not to be statistically significant, it is considered to result from random variation and is not investigated further.

6.6 Experiment 4

This experiment aims to measure shifts in the perception of non-speech sounds caused by filtered precursors. If compensation is found, Experiment 5 will be conducted to determine whether any of the mechanisms of this compensation are the same as those behind extrinsic compensation with speech. i.e. whether the enhancement effect and the spectral compensation effect explain this compensation.

6.6.1 Aims

The main aim of this experiment is to measure compensation. Listeners must compare a range of comparison sounds varying in HF content to a midpoint standard. Filtered precursors are inserted between the standard and the comparison sounds, which are expected to cause shifts in a contrastive direction to the precursor and which are indicative of compensation for the prior LTAS.

A second aim of this test is to determine whether the act of inserting precursors (regardless of their filtering) affects listener sensitivity to differences between comparison sounds and the standard, and also whether this affects the amount of compensation that occurs or can be measured. This is necessary to determine whether shifts caused by precursors cannot be measured fully due to reduced sensitivity and also to determine whether inserting a precursor reduces sensitivity so that at least 70%
sensitivity cannot be reached at continuum endpoints.

A third aim is to determine whether the act of inserting precursors causes any bias in the perception of the continuum (i.e. shifts not to do with the filtering to the precursor) and whether this affects the extent to which compensation occurs or can be measured. As stated in the pilot tests bias represents an extraneous factor affecting perception within the test and may prevent the accurate measure of shifts caused by compensation.

### 6.6.2 Hypotheses

The main hypothesis tested is that described in Section 6.2: any spectral magnitude region reduced in level by filtering to a precursor sound will be enhanced in the following comparison sound, causing a shift in perception of the spectrum of a comparison sound in the opposite direction to the filtering to the precursor.

Further, it is hypothesised that the insertion of the precursors will reduce sensitivity compared to that seen with the same programme item in the pilot tests, due to increased difficulty in comparing the standard to the comparison because of memory interference (Cowan 1984). It is also hypothesised that there will be no bias in the perception of the continuum because the stimuli range around a neutral unfiltered midpoint.

### 6.6.3 Differences between this test and the pilots

The stimuli and procedure in this test are largely the same as those described in the pilot experiments. However, there are differences that warrant laying out a description of the specific stimulus production and procedure for this experiment. Key differences between Experiment 4 and the pilot experiments are:

1. Listeners must judge differences between standard and comparison sounds with a precursor inserted between them. This is expected to a) cause shifts in the perception of comparison sounds depending on the filtering to the precursor and b) reduce sensitivity to differences between the standard and comparison sounds
compared to that seen in pilots.

2. In addition to the music-with-vocals stimulus and the noise stimulus (noise with the LTAS of the music stimulus) tested in the pilots, an instrumental item is tested to determine whether any differences between the music with and without vocals occurs due to a speech-like component in the music-with-vocals item.

3. The lengths of all stimuli are controlled in this test and timing of presentation is controlled. This is primarily to ensure that the length of time for which the precursor is presented on each trial is fixed.

### 6.6.4 Material

The music-with-vocals programme item is the same as described in Experiment Pilot 1 (see Section 6.4.2). The noise programme item is the same noise with the LTAS of the music segment described in Pilot 2 (see Section 6.5.2). A 20 ms onset/offset ramp was applied to aid smooth transitions between the previous sound or silence and each excerpt.

To create the instrumental item, an excerpt was taken from the introduction of the Jennifer Warnes Bird on a Wire pop track. This excerpt did not contain vocals. The extract was almost identical in instrumental content and temporal characteristics to the music-with-vocals excerpt and had similar spectral content. The extract was converted to mono and truncated to 5.1 seconds. As mentioned in Section 6.4.2, this length was chosen to encompass a musical phrase and ensure that it was long enough to elicit enhancement and spectral compensation mechanisms when used as a precursor.

The method used to create the standard and the HF continuum for all programme items is similar to that described for the pilot tests in Sections 6.4.3 and 6.5.3.

### 6.6.5 The standard

A standard sound was produced to act as a reference against which listeners compared sounds for HF content. To create the standard the excerpt was left unfiltered to create
a 0 dB gain natural standard sound; however, in this experiment it was decided that a roving procedure would be used whereby all sounds in the test were shifted by ± 2 dB randomly on different trials (see Table 6.10). This process creates variation in the standard. It was done to encourage subjects to listen to the standard on each trial and use it in making comparisons, rather than ignoring the standard because it was the same on each trial (or using a longer-term memory for a single standard). Listening without the standard/using a memory for the standard may cause problems with stability of ratings and may involve a different listening process that may be more or less subject to compensation. As a result a 2 dB standard was created and S/C pairs were formed using this standard as well as the 0 dB standard.

### 6.6.6 The comparison sound continuum

A number of comparison sounds were created to make the HF continuum. The comparison sounds were created by altering the 5.1 s excerpt spectrally using a high frequency shelf filter with cut off 1 kHz. To create the continuum endpoints HF above 1 kHz was boosted by -3 dB and +3 dB. Intermediate continuum points were created by altering gain above 1 kHz in 0.2 dB steps up to ± 2 dB gain and 0.5 dB steps thereafter (see Table 6.10) for the gain applied to create each continuum sound. This continuum resulted in 26 different comparison sounds. Each of the 26 comparison sounds was paired with the standard to form 26 S/C pairs (see Table 6.10), and 13 unique S/C intervals.

Pre-experiment listening by the experimenter revealed that some of the intervals could be removed to result in a shorter test (marked ‘-’ in Table 6.10). The 0.2 dB and 0.4 dB S/C intervals were removed because it was expected that these would not provide data that was more informative than at the 0, 0.6 and 0.8 dB points. It was desirable to have good resolution around the middle of the scale in order to measure JNDs at high resolution and to examine compensation here, so no stimuli were removed from this part of the continuum. It was necessary to include the 3 dB endpoint stimuli. The 2.5 dB S/C interval stimulus was also removed. This resulted in 20 different S/C pairs and (10 S/C intervals) that were used in the test. For the Noise programme item no points were removed to create a high resolution measure of this stimulus (particularly
around the 0 dB point). Therefore, there were 26 S/C pairs for the noise item and 13 S/C intervals.

Half of the S/C pairs were composed using the 2 dB standard and comparison stimuli that varied ±3 dB around this standard (i.e. from -1 dB to +5 dB, see Table 6.10). This allowed for stimulus roving whilst maintaining S/C intervals. Pre-experiment listening by the experimenter showed this continuum ranged perceptually from ‘Less’ to ‘More’ HF than the standard. During the test each S/C pair was rated once by each participant except the 0 dB pair, which was rated twice (i.e. each S/C interval was rated twice except the 0 dB interval which was rated 4 times). Because there were two precursor conditions the complete continuum was judged twice by each participant—in the ‘More’ and ‘Less’ HF than the standard precursor conditions. It was expected that overall (across precursor conditions) 50% ‘More’ HF than the standard responses would be reported by listeners judging the midpoint (0 dB) S/C interval. It was expected that near 0 and 100% ‘More’ HF than the standard responses would be expected at the -3 dB gain and +3 dB gain continuum endpoints respectively.
Table 6.10: Gain values (dB) for filtering above 1kHz. S/C pairs are formed between each comparison and the standard in the row below for the musical programme items.

<table>
<thead>
<tr>
<th>'S/C interval' Absolute gain difference (dB)</th>
<th>0.0</th>
<th>0.2</th>
<th>0.4</th>
<th>0.6</th>
<th>0.8</th>
<th>1.0</th>
<th>1.2</th>
<th>1.4</th>
<th>1.6</th>
<th>1.8</th>
<th>2.0</th>
<th>2.5</th>
<th>3.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boost</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A Comparison</td>
<td>2.0</td>
<td>-</td>
<td>-</td>
<td>2.6</td>
<td>0.8</td>
<td>3.0</td>
<td>1.2</td>
<td>3.4</td>
<td>1.6</td>
<td>3.8</td>
<td>2.0</td>
<td>-</td>
<td>5.0</td>
</tr>
<tr>
<td>B Standard</td>
<td>2.0</td>
<td>-</td>
<td>-</td>
<td>2.0</td>
<td>0.0</td>
<td>2.0</td>
<td>0.0</td>
<td>2.0</td>
<td>0.0</td>
<td>2.0</td>
<td>0.0</td>
<td>-</td>
<td>2.0</td>
</tr>
<tr>
<td>Cut</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C Comparison</td>
<td>0.0</td>
<td>-</td>
<td>-</td>
<td>1.4</td>
<td>-0.8</td>
<td>1.0</td>
<td>-1.2</td>
<td>0.6</td>
<td>-1.6</td>
<td>0.2</td>
<td>-2.0</td>
<td>-</td>
<td>-1.0</td>
</tr>
<tr>
<td>D Standard</td>
<td>0.0</td>
<td>-</td>
<td>-</td>
<td>2.0</td>
<td>0.0</td>
<td>2.0</td>
<td>0.0</td>
<td>2.0</td>
<td>0.0</td>
<td>2.0</td>
<td>0.0</td>
<td>-</td>
<td>2.0</td>
</tr>
</tbody>
</table>
6.6.7 Precursors

Listeners were presented with each S/C pair in random order across the test and had to discriminate the comparison from the standard in each pair. In all cases this comparison was made with precursors inserted between the standard and comparison sound. On each trial the S/C pair was presented with 1 of 2 precursor types (selected randomly). The precursors either contained ‘More’ or ‘Less’ HF than the standard. Precursors were made from the original 5.1s excerpt by applying either +3 dB (‘More’ HF precursor) or -3 dB (‘Less’ HF precursor) gain above 1kHz for stimuli paired with 0 dB standards, or -1 dB or +5 dB gain above 1kHz for stimuli paired with 2 dB standards. The precursors were therefore the same stimuli as those used at continuum endpoints.

This experiment aimed to produce conditions conducive to obtaining large compensation effects. According to the literature, shifts are likely to be larger with precursors that contrast highly with comparison sounds and longer precursors may cause larger compensation effects. The precursors were chosen to have the same spectra as the endpoint stimuli and represent the sound with the most and least HF content used in the test. It is noted that this is not the most extreme range of spectral colouration possible for precursors and may not result in maximal shifts. However, this choice was based on the precursors in Watkins et al’s work (Watkins 1991, Watkins and Makin 1994, 1996b), which were also stimuli from continuum endpoints. It was decided to use endpoint stimuli in this test to allow for comparison with Watkins et al’s results and also because more extreme precursors may cause unexpected results.

The precursor length was also chosen to maximise the compensation effect. Based on suggestions by Watkins (1991) and Holt (2006) longer precursors may be more likely to bring about a larger compensation effect. It is expected that sentence-length precursors may elicit larger category boundary shifts than shorter precursors (see Section 5.2). The 5.1s precursors used in the current experiments are approximately sentence length but slightly longer than those tested in previous work (specifically these are musical-phrase length). It is expected that compensation may be increased compared to previous experiments with these longer precursors. The current experiments also use the same programme items to compose the standard, precursor and comparison sound on any given trial. This helps minimise the potential for ungrouping caused by programme
item change between precursor and test sounds, which might explain the lack of compensation in some of the previous studies described in the literature review (e.g. in Watkins (1991)).

6.6.8 Method

The experimental task is similar to that described for the pilot tests. As in the pilot experiments, in each experimental session the listener was seated at a laptop computer within an ITU-R BS 1116 listening room and presented with a page-based experiment interface created in MaxMSP v 4.6. During the test the mono recordings were played diotically to listeners over Sennheiser HD 600 circumaural, open-back headphones. At the beginning of the test listeners were requested to press the play button on the computer interface to begin a sequence of trials. During each trial a new experimental page was presented. On each page an S/C pair was presented. The listener had to compare the standard to the comparison sound and determine whether the comparison had ‘More’ or ‘Less’ HF content than the standard. Unlike in the pilot tests a precursor was presented between the standard and comparison sounds and the stimulus presentation was controlled. On each trial the following sequence of sounds was presented to the listener:

1. The standard sound was heard. This lasted 5.1 s. During this time a text message was presented on screen telling participants to ‘listen’ to the spectrum of the sound.

2. There was a 0.9 s pause.

3. A 5.1 s precursor was heard, during which time a text message told listeners ‘ignore’ this sound.

4. Immediately afterwards, a 5.1 s comparison sound was heard, during which time a text message informed listeners that they should ‘compare’ this sound to the standard for spectral content.

5. A 10 s Inter-Trial Interval (ITI) occurred, during which ‘wait’ was displayed
6. The next experimental page began automatically after the ITI and the sequence was repeated.

The pause between the standard and the precursor aimed to separate the two sounds conceptually. There was no pause between precursor and comparison to ensure minimal recovery of compensation. The 10 s wait period was used to separate trials conceptually and prevent the comparison sound affecting the perception of the standard on the next trial.

On each trial listeners were asked to indicate whether the comparison sound had ‘More’ or ‘Less’ high frequency content than the standard by pressing a ‘More’ or a ‘Less’ button at any point during the ‘compare’ and ‘wait’ periods. The next trial involved the presentation of the same sequence but with new ‘standard’, ‘precursor’ and ‘comparison’ sounds. The S/C pair presented on each page was randomised, as was the presentation of the ‘More’ or ‘Less’ HF precursors with the limit that each S/C pair appeared only once in the test with each precursor type (presentation was on a Latin Squares basis). Listeners were instructed to listen for the overall spectral tilt of the sound when judging the stimuli, not just a particular frequency band. They were told that stimuli may differ by only a small amount, which they might not be able to hear, but that they should guess if they did not know the correct response. This instruction encouraged listeners to respond at chance for the 0 dB S/C intervals. Listeners were also informed that if they failed to make a response then this was noted as a missed trial and removed from analysis.

6.6.9 Subjects

The participant selection criteria are the same as described in Section 6.4.4. A larger number of participants than used in the pilots were recruited to ensure more accurate measurements. Increasing the number of participants is desirable to ensure estimates at all points are more accurate and to ensure that the psychometric functions are smoother at all points. Thirty-two subjects were tested with the music stimuli, eleven were tested with the instrumental stimuli, fifteen were tested with the noise item.
6.6.10 Music-with-vocals results

Listeners reported that they perceived sounds to vary in HF around the standard. The perception of HF content for the music-with-vocals continuum, with +3 dB (‘More HF’) precursors and -3 dB (‘Less HF’) precursors is shown in Figure 6.5. The proportion of ‘More’ HF compared to the standard responses (y axis) is plotted against the continuum points (x axis) for each precursor condition +3 dB (blue) and -3 dB (red). The green function shows the average response across the two precursors and represents listening with a neutral (0 dB) precursor. 95% confidence intervals are not displayed, due to the lack of value of these in ascertaining the significance of the difference between conditions. Instead the 95% confidence interval of the differences are displayed in Figure 6.6, presented below. The expected monotonic trend of few ‘More’ HF responses at the -3 dB continuum endpoint to many ‘More’ HF responses at the +3 dB endpoint occurred. Nearly full psychometric functions were seen with highly accurate responses at continuum endpoints for both precursor conditions (93% across both continuum endpoints and both conditions). The HF continuum was validated.

![Figure 6.5: The proportion of ‘More’ HF than standard responses for each S/C pair for the music-with-vocals programme item. Functions are plotted for the ‘More HF’ precursors (blue) and the ‘Less HF’ precursors (red) and the mean of the two (green).](image-url)
Baseline sensitivity and bias

The data of primary interest to this study are the shifts caused by precursors with different filtering. This is measured by determining the extent to which the continuum shifts in the opposite direction of the filtering to the precursors (the shift of the blue and red functions) and by examining CB measures. However, the extent to which sensitivity is reduced or bias occurs due to the act of inserting a precursor (regardless of precursor filtering) should also be assessed. This is because this baseline sensitivity and bias may affect the extent to which timbral shifts indicative of compensation occur or can be measured.

Sensitivity

The green function represents the effect of an acoustically similar but spectrally neutral sound inserted between the standard and comparison on the listeners’ sensitivity to differences between the standard and comparison sounds (it was noted in the pilot test that it is not possible to test the effect of a real neutral precursor as this would result in a repetition of the standard and this would not cause memory interference but would probably aid comparisons). The green function can therefore be examined for a change in sensitivity compared to the pilot test. In particular it should be ensured that sensitivity is not below 70% accuracy. There does not appear to be a reduction in sensitivity due to the act of inserting a precursor between the standard and comparison sound. Accuracy at continuum endpoints remains high (93% compared to 94% in the pilot test). Therefore, the data in this test can be analysed for compensation effects and compensation can be compared with that seen in other tests at low and high (at least 70%) points if desired.

Pilot test

A pilot test was not previously conducted to determine whether the ±3 dB continuum is sufficient to measure sensitivity at at least 70% accuracy in any future tests with music-with-vocals stimuli with precursors inserted between the standard and comparison sounds. The results in this test appear to show that adding a precursor of 5.1 seconds between standard and comparison does not indicate that sensitivity will be reduced in future tests with this continuum below the minimum 70% desired,
but high accuracy seen in the current test may be due to sampling variation and not representative of listening under these conditions generally. Using the current data a pilot test can be conducted to confirm that a ±3 dB continuum will be sufficient to measure compensation at high and low sensitivity with the music-with-vocals item, when a precursor is present in future tests.

Endpoint accuracy is 93% (Adjusted Wald 95% CI = 83-97%). Sampling variation may have resulted in higher accuracy than the true population accuracy in this test. Given the 93% accuracy reported in the current test a non-inferiority test shows that with 95% confidence the true population value is not below 85% ($Delta = 69\%$, $1 - 2\alpha CI (N = 64) = 85 - 96\%$) (Adjusted Wald). Given a true population value of not less than 85% it can be concluded that future tests with this continuum, under the same conditions, will not yield values lower than 70% with at least 95% confidence as long as at least 35 subjects are tested ($Delta = 69\%$, $1 - 2\alpha CI (N = 35) = 70 - 91\%$). Therefore, as with the previous test, it can be concluded that when the music-with-vocals continuum is compared to the standard binaurally with a precursor of 5.1 seconds placed between the sounds, under the same conditions seen here, accuracy is not expected to decrease to the extent that less than 70% is likely to be achieved in future tests.

However, future tests are unlikely to involve the same conditions. Noise may be presented instead of music (see equivalent analysis for noise in Section 6.6.12) or speech or instrumental music may be presented. These alternative programme items may result in lower sensitivity when listening with a precursor. However, as was explained in Section 6.4.6, it is not expected that sensitivity with these alternative programme items will reduce below that seen with music-with-vocals. An interaction between alternative programme items and listening with a precursor, such that sensitivity decreases more with a precursor for these items compared to music-with-vocals, is not expected. Therefore, as was concluded in Section 6.4.6, the current results with music-with-vocals are expected to show a conservative estimate of the sensitivity that might be obtained with different programme items. Likewise, future tests may involve a separation of stimuli in time in order to measure onset and recovery periods. As discussed in Section 6.4.6, sensitivity is not expected to reduce with less than a 30 s separation. An interaction with listening with a precursor and time gap is not expected so the minimum sensitivity for the music-with-vocals item measured here
also represents the minimum sensitivity expected with up to 30 s time gaps.

**Bias**

No shifts in the spectrum away from the continuum midpoint are expected to occur where there is no precursor or where the precursor is unfiltered. Any such shift represents a bias in the perception of the continuum. It is evident that there is a general shift in the green function towards the left hand side of the x axis in Figure 6.5. This suggests that listeners are biased towards hearing the comparison sounds as containing more HF than the standard. CB measures presented in Table 6.5 also describe this bias; CB% (overall) was 63%. This is significantly different to the expected 50% ($z = 4.9$, $p < .001$) (2-tailed). This is evidence of significant bias towards hearing the comparison sounds as containing more HF than the standard. At CB%(0 dB) listeners responded with ‘More’ 69% of the time. This is a bias of 19% from the expected 50% at this point ($z = 3.1$, $p < .002$) (2-tailed). JNDs also show bias. There is an asymmetry between the JNDs at either side of the continuum. The ‘More HF than the standard’ JND is approximately 0.2 dB and the ‘Less HF than the standard’ JND is -1.8 dB. This indicates that the listeners were biased towards making more ‘More’ HF responses across the continuum. These results show a genuine bias in the perception of the continuum, which may affect the ability to measure shifts caused by ‘Less’ HF precursors at the ‘More HF’ end of the continuum as ceiling effects may occur more readily where perception is already shifted towards hearing stimuli as containing more HF than the standard. Because the bias is significant and unexpected, attempts should be made to explain and remove this bias from future tests if possible.

It is not clear why this bias has occurred. Pilot 1 showed significant bias in the opposite direction and steps were taken to remove this by using a natural midpoint standard in the current test. However, the bias in this test may not have been completely removed as the roving procedure meant that stimuli did not always vary around a natural midpoint but a +2 dB midpoint which, as in Pilot 1, which means that half of the stimuli in the test had more HF than the original excerpt (there was an average shift of 1 dB across the whole test). As in Pilot 1 the fact that some sounds are heard to have more HF than the listener expects for this programme item may cause the bias towards more ‘More’ HF responses. However, the bias in the current test would be expected
to be in the same direction as in Pilot 1 as both tests use unnatural continua shifted in the same spectral direction (towards more ‘More’ HF responses). In the current test the bias is in the opposite direction to that in Pilot 1, so the same explanation is unlikely.

The change in the direction of the bias may be due to the addition of the precursor but it is not clear why this would result in listeners hearing the stimuli to contain more HF compared to the standard. Another explanation may come from the fact that, in the current test, listeners reported that stimuli onsets sounded particularly high in HF content (there was no report of this in the pilot test as stimuli did not have the same spectrum at onset in that test). If listeners are particularly sensitive to onset this may result in a tendency for them to hear all stimuli in the test as having more HF than expected for this programme item (particularly if they have heard this item before). As was argued for Pilot 1 this shift from the natural spectrum of the item may explain the bias. However, it would be expected that the standard would also be heard to have increased HF content and perception would be shifted in the same manner for this. Only if the comparison sounds were heard to have more HF (perhaps due to the prominence of the onset) but this was not the case for the standard (perhaps due to the onset being less prominent due to its separation in time from the judgement point) would onset perception explain the shift. Pike et al. (2014) show that onset perception in relation to a comparison sound is less prominent when the sound is separated in time from the standard. Their findings may support this conclusion.

Any baseline bias may create a ceiling effect when measuring compensation. It is concluded that attempts to remove the bias should be made when using this continuum in future tests. This may include the use of a roving standard about a neutral midpoint. However, in the current test the baseline bias is not so large as to prevent compensation being measured for low and high sensitivity at both ends of the continuum, so an analysis of compensation can proceed. Any ceiling effects due to bias should be considered in the interpretation of any compensation effects seen.
Table 6.11: Bias measures for the HF continuum composed of the Music-with-vocals programme item

<table>
<thead>
<tr>
<th>Measure</th>
<th>Expected</th>
<th>Result</th>
<th>Bias?</th>
<th>Significance of difference (2 tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CB% (overall) Proportion of ‘More’ responses across the whole continuum</td>
<td>50%</td>
<td>63%</td>
<td>Bias towards more ‘More’ responses</td>
<td>$Z = 4.9, p &lt; .001$</td>
</tr>
<tr>
<td>CB% (midpoint - 0 dB) Proportion of ‘More’ for 0 dB comparison sound.</td>
<td>50%</td>
<td>69%</td>
<td>Bias towards more ‘More’ responses</td>
<td>$Z = 3.1, p = .002$</td>
</tr>
<tr>
<td>CBdB (50%) The location of 50% ‘More’ HF responding</td>
<td>0 dB</td>
<td>Approximately - 0.8 dB</td>
<td>Bias towards more ‘More’ responses</td>
<td>-</td>
</tr>
<tr>
<td>CBdB (25%) location of 25% ‘More’ HF responding (the ‘Less HF than standard’ JND)</td>
<td>Symmetrical with CBdB (75%)</td>
<td>Approximately - 1.8 dB</td>
<td>Bias towards more ‘More’ responses</td>
<td>-</td>
</tr>
<tr>
<td>CBdB (75%) location of 75% ‘More’ HF responding (the ‘More HF than standard’ JND)</td>
<td>Symmetrical with CBdB (25%)</td>
<td>Approximately - 0.2 dB</td>
<td>Bias towards more ‘More’ responses</td>
<td>-</td>
</tr>
</tbody>
</table>
Shifts caused by precursors—0 dB continuum point

The analysis method for determining shifts indicative of compensation follows that set out in Watkins (1991). Shifts of the continuum are compared between precursor conditions at various CB points to examine compensation. Figure 6.5 shows that precursors cause consistent shifts in the perceived HF continuum in a direction which shows compensation. The shift occurs at almost all continuum points and appears larger at the continuum midpoints and smaller at continuum endpoints. The size of this shift in perception was measured by subtracting the proportion of ‘More’ responses, at the 0 dB point for the ‘More’ precursor from the proportion of ‘More’ responses, for the ‘Less’ precursor and the significance of this shift is tested. This gives a CB%(0 dB) category boundary shift of 29%. This shift is highly significant ($z = 5.0$, $p < .001$) (2-tailed). It can be concluded that there is significant compensation for the spectrum of the precursor which is demonstrated by shifts in the perceived spectrum of test sounds in the opposite direction to the filtering to the precursor.

Shifts caused by precursors—Overall measure

A measure of the shift across all S/C intervals (CB%(Overall)) was obtained. The size of the shift overall was 18%. The statistical significance of this difference was tested using a two-proportion z-test, $z = 7.0$, $p < .001$ (2-tailed). This gives the best measure of the overall effect of precursors as it measures bias at all continuum points and has the highest statistical power. It can be concluded that the shift with the music-with-vocals programme item is approximately 18% across all S/C intervals but is be smaller or non-existent where S/C intervals are large and stimuli are easy to discriminate and larger where S/C intervals are small and stimuli are more difficult to discriminate. Figure 6.11 in Section 6.6.13 shows this overall value alongside that for the instrumental music and noise stimuli.
Shifts caused by precursors—Other continuum points

An equivalent measure is made at all continuum points to obtain the difference between ‘More’ and ‘Less’ HF precursors at each individual point along the continuum. This is shown in Figure 6.6 with 95% confidence intervals for the difference. All intervals for which the confidence interval does not contain zero show a significant shift at the \( p < 0.05 \) level. The large number of these intervals not crossing zero (greater than the 1 in 20 expected by chance) shows the significance of the overall shift.

![Figure 6.6](image.png)

*Figure 6.6:* The difference in the proportion of ‘More’ HF than standard responses with ‘More HF’ and ‘Less HF’ precursors, for the music-with-vocals stimuli.

There are notable peaks and troughs in effect size across the continuum. There is a 40% effect at the -1.4 dB continuum point and -3% and 3% effect at 0.8 dB and 1 dB continuum points, respectively. These peaks and troughs may represent boundaries in the hearing system but they are not outside of those expected by random variation according to 95% CIs. There is a trend of general asymmetry in the effect size across the continuum. There appears to be a larger shift where the S/C interval is negative compared to positive. This asymmetry appears to be in part due to a ceiling effect where shifts cannot become larger at the ‘More’ HF end of the continuum due to high performance accuracy caused by the bias toward hearing stimuli as containing more HF than the standard. There are clear ceiling effects at continuum endpoints.
6.6.11 Instrumental results

The perception of HF content for the instrumental continuum, with +3 dB (More HF than the standard) precursors and -3 dB (Less HF than the standard) precursors is shown in Figure 6.7. The proportion of ‘More’ HF compared to the standard responses (y axis) is plotted against the continuum points (x axis), for each precursor condition +3 dB (blue) and -3 dB (red). The green function shows the average between the two precursors and represents listening with a neutral precursor. 95% CIs for the difference can be seen in Figure 6.8. The expected monotonic trend of fewer ‘More’ responses at the -3 dB continuum endpoint to many ‘More’ responses at the +3 dB endpoint occurred. Full psychometric functions were seen with highly accurate responses at continuum endpoints for both precursor conditions (100% across both endpoints and both conditions).

Baseline sensitivity and bias

The data of primary interest to this study are the shifts caused by precursors with different filtering. However, the extent to which sensitivity is reduced or bias occurs due to the act of inserting a precursor (regardless of precursor filtering) is also of interest. This should also be assessed as this may affect the extent to which shifts indicative of compensation occur or can be measured.

Sensitivity

The green function represents the effect of an acoustically similar sound between the comparison and the standard sounds on a listener’s sensitivity to differences between the standard and comparison sounds. For the music-with-vocals item the green function was examined for a change in sensitivity with the insertion of precursors, compared to the pilot test. As there is no pilot for the instrumental item this comparison cannot be made but sensitivity can be assessed with this item. As expected (see Section 6.4.6) sensitivity is not higher than with noise and not lower than with music-with-vocals. A continuum ranging from -3 to +3 dB was sufficient to show 100% accuracy in the current test. It can be concluded that when comparing a comparison sound to a standard with a precursor inserted between the sounds, high accuracy is achieved with the instrumental
This music item in the test (above the minimum 70% required for comparisons across tests). Therefore, this data is appropriate for further analysis of compensation.

![Figure 6.7: The proportion of 'More HF than standard' responses for each S/C pair for the instrumental programme item. Functions are plotted for More HF than the standard precursors (blue), Less HF than the standard precursors (red) and the mean of the two (green).](image)

**Pilot test**

No pilot test has been conducted with the instrumental programme item to determine whether a ±3 dB continuum range is sufficient to ensure high and low sensitivity responding in future tests with this continuum. The current sensitivity data can be used to conduct a pilot test to determine whether high sensitivity performance can be expected when conducting future tests of compensation with the instrumental programme item with a precursor inserted between the standard and comparison sound, with a ±3 dB continuum.

Accuracy across continuum endpoints is 100% (Adjusted Wald 95% CI 95% CI 83-103%). Sampling variation may have resulted in a higher accuracy in this test than the true population value. Given the 100% accuracy reported here a non-inferiority test shows that with 95% confidence the true population value is not below 87% ($Delta = 69\%, \ 1 - 2\alpha \ CI \ (N = 22) = 87 - 102\%$) (Adjusted Wald). Given a true population value of not less than 87% it can be concluded that future tests with this continuum under the same conditions may yield values as low as 73% with 95% with at least 35
participants are tested ($Delta = 69\%, 1 - 2\alpha CI (N = 35) = 73 - 93\%$). Therefore, as with the previous continua tested, values below 70% are unlikely in future tests with this continuum with this programme item under these conditions. It can be concluded that when sounds from the instrumental music continua are compared to the standard binaurally, with a precursor of 5.1 seconds placed between the sounds, accuracy is not expected to decrease to the extent that the minimum 70% is not likely to be achieved in future tests.

As with previous pilot tests it is noted that future tests are unlikely to involve the same conditions. Different programme items may be tested. As predicted in Section 6.4.6, the current results show increased sensitivity with instrumental music (with a precursor) compared to music-with-vocals (with a precursor). It is not expected that noise will result in lower sensitivity when tested with a precursor (see Section 6.6.12 for the equivalent test with noise) than that shown for the music items. Further, future tests may involve a separation of stimuli in time in order to measure onset and recovery periods. As discussed in Section 6.4.6, sensitivity is not expected to reduce before 30 s. An interaction with listening with a precursor and time gap is not expected so the minimum sensitivity measured here also represents the minimum sensitivity expected with instrumental music with up to 30 s time gaps.

**Bias**

An examination of the green function shows bias in perception of this continuum. As with the music-with-vocals stimuli, this bias is toward hearing stimuli as containing more HF content than the standard. The results for each of the bias measures is shown in Table 6.12. The overall proportion of ‘More’ responses across both precursor conditions (CB% (overall)) is 60%, which is significantly higher than the expected 50% ($z = 2.2, p = 0.027$, two tailed). At the 0 dB S/C interval (CB% (0 dB)) the proportion of ‘More’ responses is 72%. This is significantly different from the expected 50% ($z = 2.1, p = 0.034$, two-tailed).

JNDs also show bias. There is an asymmetry between the JNDs at either side of the continuum. The ‘More HF than the standard’ JND is 0 dB and the ‘Less HF than the standard’ JND is -1.8 dB. This indicates that the listeners were biased towards making
more ‘More’ HF responses across the continuum. These results show a genuine bias in the perception of the continuum, which may affect the ability to measure shifts caused by reduced HF precursors at the ‘More’ HF end of the continuum. This is because ceiling effects may occur more readily where perception is already shifted towards hearing stimuli as containing more HF. Because the bias is significant and unexpected, attempts should be made to explain and remove this bias from future tests if possible.

As with the music-with-vocals item, stimulus roving may have caused this bias. Across the test there was an average of a +1\,\text{dB} shift from the natural spectrum of the programme item. Again the bias is not in the same direction as that seen in Pilot 1, but the shift to the tests sounds is, so the same naturalness explanation may not apply in this case. This bias may explained by the onset of the stimuli being heard to contain more HF content than the remainder (the instrumental segment began at the equivalent point as the vocal item). However, as suggested for the music-with-vocals item, it would be expected that the standard would also be heard to have an increased HF onset and perception would be shifted in the same manner. It therefore appears that onset perception would only explain the shift if the comparison sounds were heard to contain increased HF (perhaps due to the prominence of the onset) but this is not the case for the standard (perhaps due to the onset being less prominent for this sound, possibly due to its separation in time from the judgement point).

It is concluded that attempts to remove the bias should be made when using this continuum in future tests. This may include the use of a roving standard that roves around a neutral midpoint because the overall 1\,\text{dB} shift in HF content from the original programme item spectrum may have resulted in the bias seen. Any baseline bias may create a ceiling effect; however, the bias as seen in the current test is not so large as to prevent compensation being measured at low and high sensitivity (at least 70\% accuracy) at both ends of the continuum. Therefore, compensation can be analysed.
Table 6.12: Bias measures for the HF continuum composed of the Instrumental programme item

<table>
<thead>
<tr>
<th>Measure</th>
<th>Expected</th>
<th>Result</th>
<th>Bias?</th>
<th>Significance of difference (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CB%(overall)</td>
<td>50%</td>
<td>60%</td>
<td>Bias towards more 'More' responses</td>
<td>$Z = 2.2$, $p &lt; .027$</td>
</tr>
<tr>
<td>CB%(0 dB)</td>
<td>50%</td>
<td>72%</td>
<td>Bias towards more 'More' responses</td>
<td>$Z = 2.1$, $p = .034$</td>
</tr>
<tr>
<td>CdB(50%) location of More HF responding</td>
<td>0 dB</td>
<td>-0.6 dB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CdB(25%) location of 25% More HF responding</td>
<td>Symmetrical with CdB(75%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CdB(75%) location of 75% More HF responding</td>
<td>Symmetrical with CdB(25%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CdB(50%) location (the 'More HF than standard' JND)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CdB(75%) location (the 'Less HF than standard' JND)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
6.6 EXPERIMENT 4

Shifts caused by precursors—0 dB continuum point

Shifts of the continuum are compared between precursor conditions. Figure 6.7 shows that precursors cause consistent shifts in the expected direction to show compensation. Shifts occur at almost all continuum points and appear larger at the continuum midpoints and smaller at continuum endpoints. The difference between the precursor conditions results in a CB%(0 dB) shift of 21%. The statistical significance of this shift was measured by calculating the average proportion of ‘More’ responses at the 0 dB continuum point for the ‘More HF’ precursor and the ‘Less HF’ precursor and testing the significance this shift. This shift was significant ($z = 2.2$, $p < .029$, two-tailed).

Shifts caused by precursors—Overall measure

The average magnitude of the shift across all S/C intervals (CB%(Overall)) is 22%. This is statistically significant ($z = 4.9$, $p < 0.001$ two-tailed). This overall shift is also shown in Section 6.6.13 in Figure 6.11, alongside that for the music-with-vocals programme item and the noise programme item.

Shifts caused by precursors—Other continuum points

An equivalent measure at other continuum points is shown in Figure 6.8 with 95% confidence intervals for the difference. All intervals for which the confidence interval does not contain zero, show a significant shift. Again, it is apparent that the shift is larger where the S/C interval is small and sensitivity is low and the shift is smaller where sensitivity is high. There is asymmetry in the effect. There is up to a 45% shift when listeners are presented with comparison sounds from the negative side of the continuum (-1.4 dB and -1.2 dB less HF than the standard). There is a 0 dB shift when listeners are presented with sounds that contain 1 dB and 1.6 dB more HF than the standard. As suggested for the musical programme item, this asymmetry may be due to random variation as confidence intervals show that the shifts at these continuum points are not beyond that expected by chance. However, there may be some underlying asymmetry in perception that requires further testing. There appears to be evidence of a general
asymmetry that shows a ceiling effect where shifts cannot become larger at the ‘More’ HF end of the continuum because sensitivity is already high in this region due to the bias. However, only limited conclusions should be drawn about this asymmetry as confidence intervals show that the peak and trough are not significantly different from neighbouring data and are within the range expected from sampling variation. There are clear ceiling effects at the continuum endpoints.

![Figure 6.8: The difference in the proportion of ‘More’ HF than standard responses for each S/C pair between ‘More HF’ and ‘Less HF’ precursors, for the instrumental stimuli.](image)

### 6.6.12 Noise results

The perception of HF content for the noise continuum with +3 dB (‘More HF’) precursors and -3 dB (‘Less HF’) precursors is shown in Figure 6.9. The proportion of ‘More’ HF responses (y axis) is plotted against the continuum points (i.e. the S/C pairs) (x axis) for each precursor condition; +3 dB (blue) and -3 dB (red). The green function shows the average between the two precursors and represents listening with a neutral precursor. 95% confidence intervals for the difference are shown in Figure 6.10. The expected monotonic trend of fewer ‘More’ responses at the -3 dB continuum endpoint to many ‘More’ responses at the +3 dB endpoint occurred. Nearly full psychometric functions were seen with highly accurate responses at continuum endpoints for both precursor conditions (96.5% across both continuum endpoints and both conditions).
Baseline sensitivity and bias

The data of primary interest to this study are the shifts caused by precursors with different filtering. However, the extent to which sensitivity is reduced or bias occurs due to the act of inserting a precursor (regardless of precursor filtering) should also be assessed as this may affect the extent to which shifts indicative of compensation occur or can be measured.

Sensitivity

The green function represents the effect of an acoustically similar but spectrally neutral sound inserted between the standard and comparison on the listeners’ sensitivity to differences between the standard and comparison sounds (it was noted in the pilot test that it is not possible to test the effect of a real neutral precursor as this would result in a repetition of the standard and this would not cause memory interference but would probably aid comparisons). The green function can therefore be examined for a change in sensitivity compared to the pilot test. In particular it should be ensured that sensitivity is not below 70% accuracy. There is some reduction in sensitivity due to the act of inserting a precursor between the standard and comparison sound but accuracy at continuum endpoints remains high (96.5% compared to 100% in the pilot
test). Therefore the data in this test can be analysed for compensation effects and compensation can be compared with that seen in other tests at low and high (at least 70%) points if desired. There is slightly more reduction in accuracy due to inserting the precursor seen for noise compared to music-with-vocals (3.5% compared to 1%) (no pilot was conducted for instrumental music so this comparison cannot be made). The precursor involves a separation for the standard and comparison. This may show that with noise having the stimuli side-by-side is more important to judging accuracy than with the music items. However, low power prevents a significance test of this small interaction so no conclusions can be made.

Pilot

A pilot test was previously conducted with noise that showed high sensitivity when listening without a precursor. It was noted that listening with a precursor may result in lower sensitivity but this could not be measured in the previous pilot test. It is useful to determine whether the ±3 dB continuum is sufficient to measure sensitivity at at least 70% accuracy for noise with a precursor inserted between the standard and comparison sound in any future tests. The same analysis as presented in Pilot 1 with binaural and monaural listening without a precursor is conducted for listening with a precursor.

Endpoint accuracy is 96.5% (Adjusted Wald 95% CI 78-99%). Sampling variation may have resulted in a higher accuracy than the true population value. Given the 96.5% accuracy reported in the current test, a non-inferiority test shows that with 95% confidence the true population value is not below 81% ($\Delta = 69\%$, $1 - 2\alpha CI (N = 30) = 81 - 99\%$) (Adjusted Wald). Given a true population value of not less than 81% it can be concluded that future tests with this continuum under the same conditions may yield values lower than 70% with 95% confidence with at least 35 participants tested ($\Delta = 69\%$, $1 - 2\alpha CI (N = 35) = 67 - 89\%$). It is possible that this continua may yield values below 70% in future tests, but this is unlikely due to the conservative nature of the calculation. The conservative estimate seen is probably due to the relatively low power of this analysis and it is likely that the true population value is higher than this estimate. However, to confirm the suitability of this continuum in future tests more data should be collected and the results should be analysed again.
If the non-inferiority test with more data still shows that values less than 70% may be obtained in future tests then the continuum may need to be expanded for testing noise with a precursor inserted between the standard and comparison sound to ensure that at least 70% responding is obtained in future tests. However, as highly accurate responses were obtained in the current test, compensation effects can be analysed and compared with other tests at at least 70% accuracy if desired.

As noted previously future tests are unlikely to involve the same conditions. For example, future tests may involve a separation of stimuli in time in order to measure onset and recovery periods. Sensitivity is not expected to reduce before 30 s and, as for the other programme items, an interaction with listening with a precursor and time gap is not expected for noise. Therefore, the minimum sensitivity for the noise item measured here also represents the sensitivity expected with a precursor and a time gap up to 30 s.

**Bias**

It is evident from the green function in Figure 6.9 that there is a general shift for all continua towards the right hand side of the x axis. This shows a tendency for fewer ‘More’ HF responses and a bias towards hearing the stimuli as containing less HF compared to the standard. This shift is described by CB measurements presented in Table 6.13. However, this shift is not significant according the overall measure. CB%(Overall) is 45% compared to the expected 50% (z = 1.4, p = .162). Power is high for the overall measure so it is unlikely that low significance is explained by this. It is also not significant at the continuum midpoint. CB%(0dB) is 38% compared to 50% Z = 0.9, p = .349. The other continuum measures also show bias towards making fewer ‘More’ responses. The ‘More HF than the standard’ JND is 1.2 dB and the ‘Less HF than the standard’ JND is -0.6 dB, showing increased sensitivity at the ‘More HF’ end of the continuum and indicating a bias towards hearing the stimuli as containing more HF than the standard. It is possible that this small bias is a result of sampling variation but a further examination of the source of this bias should be made and it should be eliminated from future tests with this continuum if possible. Bias with the noise stimuli is unexpected as it cannot be explained by any difference in perception at onset. The spectra of the excerpts are uniform so onsets would not have
been heard to contain increased HF content. This finding is supported by Pike et al. (2014) who showed that listeners appear to be biased towards judging a comparison sound based on its onset against the standard with music, but not with noise (with the same excerpts as used here). Further, there is no natural spectrum for this programme item so a general shift towards the sound being perceived as containing more HF than is natural due to roving cannot explain this bias. As the bias is non-significant and there is no apparent explanation it is concluded that this bias may be an artefact of sampling variation. It is evident that bias seen in the current test is unlikely to affect the measurement of compensation and further analysis of compensation can proceed.


Table 6.13: Bias measures for the HF continuum composed of the Noise programme item

<table>
<thead>
<tr>
<th>Measure</th>
<th>Expected</th>
<th>Result</th>
<th>Bias?</th>
<th>Significance of difference (2 tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CB%(Overall) Proportion of ‘More’ responses across the whole continuum.</td>
<td>50%</td>
<td>45%</td>
<td>Bias towards fewer ‘More’ responses</td>
<td>$Z = 1.4, p &lt; .162$</td>
</tr>
<tr>
<td>CB%(0 dB) Proportion of ‘More’ 0 dB comparison sound.</td>
<td>50%</td>
<td>38%</td>
<td>Bias toward fewer ‘More’ responses</td>
<td>$Z = 0.9, p = .349$</td>
</tr>
<tr>
<td>CBdB (50%) The location of 50% ‘More’ HF responding</td>
<td>0 dB</td>
<td>0.4 dB</td>
<td>Bias toward fewer ‘More’ responses</td>
<td>-</td>
</tr>
<tr>
<td>CBdB (25%) The location of 25% ‘More’ HF responding (the ‘Less HF than standard’ JND) Symmetrical with CBdB(75%)</td>
<td>Symmetrical with CBdB(75%)</td>
<td>-0.6 dB</td>
<td>Bias towards fewer ‘More’ responses</td>
<td>-</td>
</tr>
<tr>
<td>CBdB (75%) The location of 75% ‘More’ HF responding (the ‘More than standard’ JND) Symmetrical with CBdB(25%)</td>
<td>Symmetrical with CBdB(25%)</td>
<td>1.2 dB</td>
<td>Bias towards fewer ‘More’ responses</td>
<td>-</td>
</tr>
</tbody>
</table>
6.6 EXPERIMENT 4

Shifts caused by precursors—0 dB continuum point

Shifts of the continuum are compared between precursor conditions. Figure 6.7 shows that precursors cause consistent shifts in direction expected to show compensation. The shift occurs at almost all continuum points, and is larger at the continuum midpoints and smaller at continuum endpoints. The difference between the precursor conditions results in a CB%(0 dB) shift of 30%. This difference was significant ($z = 2.4, p < .017$, two-tailed).

Shifts caused by precursors—Overall measure

When averaged across all SC intervals the shift is large. CBdB(Overall) = 26%. This shift is significant ($z = 7.3, p < .001$, two-tailed). This shift can be seen in Section 6.6.13 in Figure 6.11 alongside shifts for the other programme items.

Shifts caused by precursors—Other continuum points

Figure 6.10 shows the difference between ‘More’ and ‘Less’ HF precursors at each S/C interval with 95% confidence intervals for the difference. Few confidence intervals cross the zero line, demonstrating the overall significance of the shift. Shifts tend to be larger where the S/C interval is small. There is less evidence of asymmetry in the effect with the noise item compared to the music programme items. This supports a conclusion that this asymmetry seen in previous tests is largely due to the general shift in the continuum caused by baseline bias. There is evidence of a slight asymmetry which is the opposite to that seen in previous tests and here, which may indicate the impact of the slight baseline bias towards hearing the stimuli as containing less HF, causing ceiling effects at this end of the continuum. However, no conclusion can be drawn about the asymmetry as it not beyond that expected by the confidence intervals.
6.6 EXPERIMENT 4

Figure 6.10: The difference in the proportion of ‘More’ HF than standard responses for each S/C pair between ‘More’ and ‘Less’ HF precursors, for the noise stimuli. 95% confidence intervals for the difference are plotted.

6.6.13 Comparison of all 3 programme items

A significant shift indicative of compensation occurred for all the non-speech programme items tested. It can be concluded that compensation occurs in this entirely non-speech test as well as where speech is present in the test stimuli. It is apparent that there are differences in the sizes of shifts between programme items. When averaged across all S/C pairs (CB%(Overall)) the music-with-vocals programme item shows an 18% shift, the instrumental item shows a 22% shift, and the noise item shows a 26% shift. Figure 6.11 shows the shift for each programme item along with the 95% confidence interval. The difference between music-with-vocals and noise was not significant, nor was the difference between music-with-vocals and instrumental music. The difference between the music-with-vocals and the noise programme item is significant ($z = 3.1, p < .002$). This remains significant with the use of a Bonferroni adjusted p-value (0.006) to account for the fact that 3 comparisons are made between the programme items. This result is suggestive of compensation differences depending on the programme item. Contrary to a hypothesis that speech may elicit more compensation compared to noise, the shift is smaller for the music-with-vocals item, which contains a speech-like element. This finding suggests that speech may not benefit
from a special compensation mechanism. However, it is still possible that, regardless of the size of the shift, the shift with this item is caused by a different mechanism to that which arises with noise. A special speech mechanism is not ruled out by these findings. However, it is expected that this stimulus is heard as music rather than speech. It is also apparent that the stimuli with spectral variation do not result in a larger shift than noise. This may be evidence against the spectral compensation effect occurring in the perception of these sounds but again it is possible that regardless of size the shift seen with these items it is caused by a different mechanism to that seen with noise.

![Figure 6.11: Difference in proportion of 'More HF' responses between the 'More' and 'Less' HF precursors averaged across all S/C intervals for the music-with-vocals, instrumental and noise programme items.](image)

**6.6.14 Experiment 4 Summary and discussion**

The aim of this experiment was to measure shifts in the category boundary of a non-speech HF continuum when filtered non-speech precursors were presented before sounds from that continuum (music-with-vocals, instrumental music and noise where tested). A secondary aim was to determine whether the act of inserting precursors (regardless of their filtering) affects listener sensitivity to differences between the comparison sounds and the standard and whether this prevents compensation being measured at high and
low sensitivity points. A third aim was to determine whether there was any bias in the perception of the continua that may affect the extent to which compensation occurs or can be measured.

To address the first aim the tests determined whether shifts in the perceived spectrum of non-speech comparison sounds occurred in the opposite direction to the filtering applied to non-speech precursor sounds. The primary rationale was to determine whether such shifts occur at all, as these are evidence for compensation for transmission channels that occurs through prior listening to the channel (extrinsic compensation). In particular such shifts are evidence of specific extrinsic compensation mechanisms, the enhancement effect and the spectral compensation effect, occurring with non-speech. As shown in Section 5.4, such extrinsic compensation mechanisms may play a role in compensation seen in real-world tests and in particular may explain the time-gap sensitive component of that compensation. Exactly which of the mechanisms, the enhancement or the spectral compensation effect, cause the shifts will be determined in Experiment 5.

Shifts indicative of compensation were seen with all non-speech programme items. There was a similar size shift (18% music-with-vocals, 22% instrumental music and 26% noise) with all programme items. A 100% shift would be evidence for a complete reduction in the channel, so the compensation is shown to be small to moderate. However, small compensation may be adequate to enhance the perception of a changing sound source relative to the channel. This may be adequate to remove the channel to the extent that some spectral information about the channel is maintained, such as the average spectrum of a talker’s voice (which may be used in voice identification), whilst enhancing the perception of the source. The existence of shifts with non-speech stimuli indicates that compensation does not rely on speech-specific mechanisms. This finding concurs with those of recent studies that show shifts in the perception of vowel sounds after non-speech precursors, shifts with non-speech test sounds after speech precursors and shifts in an entirely non-speech context. The results add evidence to the claim that the extrinsic channel compensation seen in these studies is a general auditory process.

There was some difference between programme items in the magnitude of compensation. The difference in shift size between instrumental and vocal music was not
significant. This does not imply that the mechanisms of compensation are the same, as it is possible that the same size shift can arise by completely different mechanisms. Further, this result does not indicate that the shifts are not different in size, just that the current evidence is not strong enough to show they are different. However, the reasonably high power of the analysis suggests that if any real difference exists in shift size between instrumental and vocal music it is likely to be small. The difference in shift between instrumental music and noise was also not significant. Again this does not suggest that shifts are caused by the same mechanisms or that the shift is exactly the same size (but the high power of this analysis shows that any real difference is likely to be small). The shift with music-with-vocals was shown to be significantly smaller than the shift with noise so it can be concluded that the shifts are different in magnitude. This does not imply entirely different mechanisms, as it is possible that shifts of different sizes come about by the same mechanism with different programme items. Tests to determine the mechanisms of these shifts with different programme items should be conducted.

It might be expected that the music-with-vocals programme item and the instrumental item would engage a larger rather than smaller compensation effect compared to noise as these stimuli are expected to engage the spectral compensation effect, either because they contain speech-like content (music-with-vocals) or because they contain spectro-temporal variation. The fact that music-with-vocals produced the smallest shift suggests that the inclusion of a speech-like element in the track does not lead to additional compensation due to the triggering of a speech-specific perceptual mechanism. It appears that sung speech within music is not heard as speech but as music. Listeners may be more likely to engage with speech as speech when listening to plain speech in an information-bearing context, rather than when it is sung in a musical context. Therefore this result does not show that shifts with music and speech are a similar size. The similarity between instrumental and music-with-vocals items suggests that it may be possible to regard the music-with-vocals item as simply an example of music for the purposes of future tests but it remains possible that, although shifts are the same size with both items, the mechanisms behind both are different. A test of mechanisms is necessary to confirm that music with and without vocals undergo compensation via the same mechanisms.
The finding of a smaller effect with both music-with-vocals and instrumental music compared to noise suggests that spectral variation does not engage an additional effect. Noise is only expected to benefit from the enhancement effect but the music items are expected to benefit from the spectral compensation effect plus enhancement (Watkins 1991). These results may be evidence of all three items only experiencing shifts caused by the enhancement effect. If the enhancement effect is caused by simple peripheral adaptation (or adaptation of suppression) it may expected that effects would be smaller with stimuli with spectral variation compared to static stimuli (see Section 5.2). Therefore the reduced magnitude of the effect with music compared to noise may be indicative of all shifts being caused by the enhancement effect. However, the working of the spectral compensation effect is not ruled out. It could be that with noise only the enhancement effect occurs, but with the music items the enhancement effect is small (possibly due to the spectral variation) and the remainder of the shift seen is produced by the spectral compensation effect. A further possibility is that the spectral compensation effect occurs with noise as well as music and this explains the equivalence. However, this is unlikely as a number of prior studies have shown only a monaural enhancement effect with noise. The answers to these questions can only be obtained by tests examining the specific mechanism behind the effects.

Differences in magnitude of shifts between programme items do not appear to be explained by different ceiling effects. Differences between items were only compared using the CB(Overall) measure and ceiling effects are more likely to reduce the shift measured by the CB(Overall) measure with noise. This is because noise is a higher sensitivity item and the ceiling is likely to reached earlier in the continuum (i.e. on average, shifts are being measured at lower sensitivity for music than noise and so overall the shift is expected to be larger). Therefore increased ceiling effects with music cannot explain smaller shifts. Measures at individual points such as the CB%(0 dB) can also be examined as shifts are not affected by ceiling effects with this measure. However, this also does not compare shifts at equal sensitivity points (for example, bias may shift the continuum so at 0 dB listeners may make 100% ‘More’ or ‘Less’ HF responses for one programme item but not the other. Compensation would be expected to be different at 0 dB because of the higher sensitivity performance exhibited for one programme item compared to another).
A more equivalent measure of compensation may be made by comparing shifts at the same sensitivity points across programme items. For example, the shift at the 50% sensitivity point can be examined for each programme item. As was noted in Section 6.3.2, confidence intervals for this measure could not be calculated so the measure is instead translated to the continuum point that corresponds to 50% sensitivity for each programme item (CBdB(50%)) and a comparison of shifts at this point are be made. It can be seen that 50% equates to -0.8 dB on the continuum for the music-with-vocals item. At this continuum point the shift is 19%. For the instrumental item CBdB(50%) equates to -0.6 dB on the continuum. At this point the shift is 25%. For noise the 50% point is at +0.4 dB and here the shift is 53%. It is apparent that the shifts at equal sensitivity points are either as large as or larger than those measured using the CB(Overall) measure. The power is too low to compare significance at equal sensitivity points in the current tests. Compared to the 100% shift that would be expected with full compensation it may still be concluded that the shifts measured at this point, although larger, range from small to moderate. The aim of making comparisons at equal sensitivity points is to make more valid comparisons between programme items (or test conditions). In terms of a comparison between programme items the trend of larger shifts with noise compared to music remains but is exaggerated when compared at equal sensitivity points.

The ceiling effects at the continuum endpoints were apparent. Ceiling effects may either prevent shifts occurring (it may be that once differences between the standard and comparison are large perception is categorical and shifts do not occur), or they may prevent shifts being measured (it may be that once the difference is large, shifts occur but do not have an instrumental effect of labelling, i.e. the listener hears the stimulus as containing more HF but this does not have an effect on the accuracy with which they label the stimulus as they are already making 100% ‘More’ HF responses). With more trials it may be possible to observe shifts for endpoint stimuli as differences may become apparent in long-term statistics. Alternatively, a method of response that allows listeners to report a shift too small to result in a different categorisation as ‘More’ or ‘Less’ HF content than the standard may be desirable in future tests. This might be achieved by eliciting subjects’ judgements of the confidence they have in their answers. Reaction times can also be used to measure the ease with which stimuli are categorised.
The lack of shifts at stimulus endpoints is suggestive that compensation and categorical perception are related. There may be a link between the effects seen here. This shifting of perception to a larger difference between the standard and comparison sounds also appears to be similar to the shifting seen with the perceptual magnet effect that shifts perception to category prototypes. Further, as noted already, where sensitivity is greater the shift measured with the Overall measure may be reduced as more of the continuum is affected by ceiling performance, creating unequal comparisons. It may therefore be preferable to attempt to measure compensation where ceiling effects are minimised or only measure shifts at equal sensitivity points rather than with the overall measure in future tests.

6.6.15 Experiment 4 Conclusion

The current study aimed to determine whether, with non-speech stimuli, spectral filtering to a precursor sound creates a shift in the perception of the spectrum of a following comparison sound, indicative of compensation by extrinsic compensation mechanisms (the enhancement effect and the spectral compensation effect). A shift in perception of the comparison sound was observed that was approximately 26% of the range of the comparison sound continuum for noise, 22% for instrumental music and 18% with music-with-vocals when compared at equal continuum points. Shifts were larger when compared at equal sensitivity points and the pattern of larger shifts for noise compared with music-with-vocals remained. It cannot be said that differences between programme items are due to the ease with which discrimination between the standard and comparisons can be made and the ceiling effects that this may entail. The small to moderate shifts observed could indicate non-complete compensation, which may be advantageous to hearing channel effects where necessary but may be sufficient to allow more constant perception of sound sources across different channels.

A second aim of this work was to ensure that future tests with precursors do not result in sensitivity at continuum endpoints of lower than 70% as this would mean that such tests would not be able to measure compensation at high (at least 70%) sensitivity. This would also mean that it may not be possible to compare the results at a single high accuracy point between the current tests and those tests. It was shown that future tests
with the music stimuli will be expected to yield at least this high accuracy performance at continuum endpoints. However, the noise item should be tested further to ensure this. Further, an aim was to ensure that large bias was not present in the current test as this could prevent the measurement of compensation due to ceiling effects. Some bias was present but this was not large enough to prevent a measurement of compensation in this case. Where this bias was significant (for music-with-vocals and instrumental music) attempts should be made to eliminate it in future testing. This work contributes to answering Research Question 1b and Step 1a(ii) of the research process. It was shown that at least one of the extrinsic compensation mechanisms occurs with non-speech listening. This may therefore be considered a potential mechanism of compensation for loudspeakers and rooms. However, further work is necessary to determine which of these specific mechanisms is behind the compensation seen in this test. Tests to determine this will be carried out in the next subsection.

6.7 Experiment 5

Experiment 5 aimed to test whether the shifts seen in Experiment 4 occur when the precursor and comparison sounds are presented to different ears (crossaural presentation). This is to determine whether or not the mechanisms of compensation observed in Experiment 4 are crossaural. If they are not then the mechanism is likely to be peripheral and caused by the auditory enhancement effect. If the effect is crossaural then it cannot be peripheral and must be a central mechanism; such an effect is likely to be a manifestation of the spectral compensation effect.

To test the crossaural nature of the shifts a test was conducted that followed all methods and procedures used in Experiment 4, except that the standard and comparison sounds were presented to one ear and the precursor was presented to the alternate ear.

6.7.1 Participants

The selection criteria were the same as described in Section 6.4.4. Thirty-five subjects were tested with the music-with-vocal stimuli, and fourteen with the noise item.
6.7.2 Material

The music-with-vocals programme item is the same as described in Section 6.4.2 and the noise programme item is this same noise with the LTAS of the music segment as described in Section 6.5.2. The instrumental item was not tested.

6.7.3 Method

The method was the same as that described in Section 6.6.8, except that on each trial the standard and comparison sounds where presented to one ear and the precursor sounds were presented to the other.

6.7.4 Results—Music-with-vocals

There were no problems reported in making HF judgements. Listeners reported that they perceived sounds to vary in HF. The perception of HF content for the music-with-vocals continuum, with +3 dB (‘More’ HF) precursors and -3 dB (‘Less’ HF) precursors is shown in Figure 6.12. The proportion of ‘More’ HF compared to the standard responses (y axis) is plotted against the continuum points (x axis) for each precursor condition +3 dB (blue) and -3 dB (red). The green function shows the average between the two precursors and represents listening with a neutral precursor. 95% confidence intervals of the difference between the two functions can be seen in Figure 6.13. The expected monotonic trend of fewer ‘More’ responses at the -3 dB continuum endpoint to many ‘More’ responses at the +3 dB endpoint occurred. Nearly full psychometric functions were seen with highly accurate responding at continuum endpoints for both precursor conditions.

Baseline sensitivity and bias

The data of primary interest to this study are the shifts caused by precursors with different filtering. This is measured by determining the extent to which the continuum shifts in the opposite direction of the filtering to the precursors (the shift of the blue and
red functions), and by examining CB measures (see Section 6.7.4). However, the extent to which sensitivity is reduced or bias occurs due to the act of inserting a precursor (regardless of precursor filtering) should also be assessed. This is because this baseline sensitivity and bias may affect the extent shifts indicative of compensation occur or can be measured.

**Sensitivity**

Reduced sensitivity can prevent compensation being measured at high sensitivity points sensitivity should therefore be examined. The green function in Figure 6.12 represents the effect of an acoustically similar but spectrally neutral sound inserted between the standard and comparison sounds on sensitivity. It can be seen that in a test where the standard and comparison sounds are compared monaurally and where a precursor is presented contralaterally sensitivity remains high at continuum endpoints (94.5% across continuum endpoints). As expected, sensitivity was reduced somewhat compared to monaural listening without a contralateral precursor (100%, see Section 6.4.6), showing that the task of comparing stimuli monaurally with a contralateral precursor may be more difficult (statistical power at the endpoints is not large enough to test for the significance of the difference). Importantly accuracy in this test is not below 70% meaning that it is possible to analyse compensation and make
a comparison of compensation in this test and others with at least 70% sensitivity.

**Pilot test**

Future tests with crossaural precursor presentation are likely to be conducted with the same programme item in different conditions (e.g. with different time gaps between standard and comparison sounds when testing the time course of the enhancement effect, see Chapter 6). The high accuracy seen in the current test may be due to sampling variation and not representative of listening under these conditions generally. A pilot test was not previously conducted to determine whether the ±3 dB continuum is sufficient to measure sensitivity at at least 70% with monaural listening with a contralateral precursor in future tests. However, it was concluded that monaural listening without a precursor would not lower sensitivity below 70% (see Section 6.4.6). The same analysis as presented in Pilot 1 for monaural listening without a precursor is conducted here for monaural listening with a contralateral precursor.

The pooled accuracy at both endpoints gives an endpoint accuracy of 94.5%. A non-inferiority test shows that with 95% confidence the true population value is not below 84% (\( \Delta = 69\% \), \( 1 - 2\alpha CI (N = 70) = 88 - 98\% \) (Adjusted Wald). Given a true population value of not less than 88% it can be concluded that future tests with this continuum under the same conditions will yield values not lower than 73% with 95% confidence, if at least 35 participants are tested (\( \Delta = 69\% \), \( 1 - 2\alpha CI (N = 35) = 73 - 93\% \)). Therefore, as with the previous continua tested, values below 70% are unlikely to occur in future tests with this continuum and this programme item under these conditions. It can be concluded that when stimuli are compared monaurally, adding a precursor of 5.1 seconds to the contralateral ear does not decrease sensitivity to the extent that the minimum 70% is not likely to be achieved in future tests. However, future tests are unlikely to involve the same conditions. Noise may be presented instead of music (see equivalent analysis for noise in Section 6.7.5) and instrumental music may be presented. These programme items may result in lower sensitivity when listening monaurally with a contralateral precursor. However, as explained in Section 6.4.6, it is not expected that sensitivity will reduce below that seen with music-with-vocals as this is thought to represent the most difficult to discriminate item. An interaction between other programme items (e.g. instrumental music, noise) and the crossaural listening
conditions such that sensitivity decreases compared to that seen in the current test is not expected. The results with music-with-vocals show a conservative estimate of sensitivity when listening monaurally with a contralateral precursor and with different programme items. Likewise, future tests may involve a separation of stimuli in time for measuring onset and recovery periods. As discussed in Section 6.4.6, sensitivity is not expected to reduce before 30 s elapses. An interaction with monaural listening, a contralateral precursor and time gap is not expected. Again, the minimum sensitivity for the music-with-vocals item measured here also represents the minimum sensitivity expected with a time gap up to 30 s.

Bias

It is evident from the green function in Figure 6.12 that there is a general shift towards the left hand side of the x axis. This shift is described by CB measurement presented in Table 6.15. At CB% (0 dB) listeners responded ‘More’ 62% of the time. This shift shows bias towards hearing the continuum as containing more HF content than the standard. This is a shift of 12% from the expected 50% point. A two-proportion z-test showed this to be significantly different from the expected 50% ‘More’ responses ($z = 2.0, p < .043$). The bias across all S/C intervals (Overall bias) was calculated by the proportion of ‘More’ responses across the whole test. This was 59% and so was also significantly different to the expected 50%, ($z = 3.5, p < .001$). The bias is most apparent when looking at the asymmetry in JND measures. The ‘Less HF than the standard’ JND is 2 dB but the ‘More HF than the standard’ JND is 0.8 dB. This indicates strong bias in favour of hearing stimuli as containing more HF compared to the standard.

It is not clear why this bias has occurred. Pilot 1 showed bias in the opposite direction, towards hearing the stimuli as containing less HF than the standard. However, in tests with music-with-vocals and instrumental music where a precursor was inserted and the stimuli length controlled for (Experiment 4) there was bias towards hearing the stimuli to contain more HF than the standard. The results in the current study may be further evidence that insertion of the precursor has an effect on bias toward hearing stimuli as containing more HF than the standard. It may also be further evidence that the onsets of the stimuli being perceived to contain more HF content may cause this bias and that
when judgements are made the comparison sound’s onset is more prevalent than the onset of the standard. If this is the explanation for bias it is apparent that placing the precursor in the contralateral ear does not prevent this bias. The bias may also be caused by of the overall 1 dB shift of the sounds in this test away from the natural spectrum of the music item due to stimulus roving. Future tests with this continuum should aim to take steps to remove this bias. However, it is apparent that the bias here is not likely to prevent the measure of compensation at high and low sensitivity points.
Table 6.14: Bias measures for the HF continuum composed of the music-with-vocals programme item with monaural comparisons and a precursor presented contralaterally.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Expected</th>
<th>Result</th>
<th>Bias?</th>
<th>Significance of difference (2 tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CB%(overall) Proportion of ‘More’ responses across the whole continuum</td>
<td>50%</td>
<td>59%</td>
<td>Bias towards more ‘More’ responses</td>
<td>$Z = 3.5, p &lt; .001$</td>
</tr>
<tr>
<td>CB%(0 dB) Proportion of ‘More’ 0 dB comparison sound.</td>
<td>50%</td>
<td>62%</td>
<td>Bias toward more ‘More’ responses</td>
<td>$Z = 2, p = .043$</td>
</tr>
<tr>
<td>CBdB(50%) The location of 50% ‘More’ HF responding</td>
<td>0 dB</td>
<td>-0.8 dB</td>
<td>Bias toward more ‘More’ responses</td>
<td>-</td>
</tr>
<tr>
<td>CBdB(25%) The location of 25% ‘More’ HF responding (the ‘Less than the standard’ JND)</td>
<td>Symmetrical CBdB(75%)</td>
<td>-2 dB</td>
<td>Bias towards more ‘More’ responses</td>
<td>-</td>
</tr>
<tr>
<td>CBdB(75%) The location of 75% more HF responding (the ‘More than the standard’ JND)</td>
<td>Symmetrical with CBdB(25%)</td>
<td>0.8 dB</td>
<td>Bias towards more ‘More’ responses</td>
<td>-</td>
</tr>
</tbody>
</table>
Shifts caused by precursors—0 dB continuum point

Shifts of the continuum are compared between precursor conditions. Figure 6.12 shows that precursors cause consistent shifts in the expected direction for compensation. The shift occurs at almost all continuum points. Shifts appear largest to the left of the continuum due to a general shift of the continuum to the left. The size of compensation was measured by subtracting the proportion of ‘More’ responses at the 0 dB point for the ‘More’ HF precursor from the proportion of ‘More’ responses for the ‘Less’ HF precursor. This gives a non-significant CB% (0 dB) boundary shift of 5%. ($z = 0.9$, $p < .390$, two-tailed). However, it is apparent that this lack of shift is not representative of that seen in the rest of the continuum and may due to the bias and may represent sampling variation.

Shifts caused by precursors—Overall measure

A measure of the shift across all S/C intervals, is also obtained. The size of the shift overall was 8%. The statistical significance of this difference was tested using a two-proportion z-test. The shift was highly significant ($z = 3.2$, $p < .001$, two-tailed). It can be concluded that the shift with the music-with-vocals programme item, when the precursor is presented crossaurally, is approximately 8% across all S/C intervals. This represents a highly significant effect. However, the effect is not evident at some continuum points.

Shifts caused by precursors—Other continuum points

The shift is smaller at continuum endpoints. As previously noted, high sensitivity performance appears to have prevented compensation from occurring or being measured at these points. These ceiling effects appear to have prevented a larger overall shift from being observed. A measure of the shift at each continuum point is obtained. This is shown in Figure 6.13 with 95% confidence intervals for the difference. All intervals for which the confidence interval does not contain zero show a significant shift at the $p < 0.05$ level.
The results show evidence of a small consistent shift. The shift is in the same direction in all cases. Three of the confidence intervals do not cross zero which is more than expected for this many comparisons (1 in 20 are expected to not cross zero by chance). A line can be drawn which includes 95% of the confidence intervals and this is non-zero which demonstrates the significance of the overall shift. It should be noted that this analysis represents a two-tailed test, whereas it would also be appropriate to perform a one-tailed test on this data, as the direction of the shift was predicted before collecting data and the shift is in the expected direction. Given this, 90% confidence intervals would show significant shift even more clearly. Asymmetry of the effect size across the continuum is not obvious, but shifts appear larger at the negative end of the continuum which is expected due to the nature of the bias seen.

**Music-with-vocals reduction in effect**

A shift was seen with the music-with-vocals item when the precursor was presented contralaterally. This reduced from the 18% shift in Experiment 4 with binaural presentation of the precursor. The reduction of 10% from the shift seen in Experiment 4 with the same programme item may indicate a role for a peripheral or ear-sensitive
mechanism when listening to music-with-vocals. A comparison of the overall measures shows that the reduction in shift between Experiment 4 and 5 was significant ($z = 5.7$, $p < .001$). The remaining shift is also significant ($z = 3.2$, $p < .001$). The shift remaining shows a central contribution to the compensation seen in Experiment 4.

A comparison is also made at equal sensitivity points. For the ipsilateral precursors condition there was a shift at CBdB(50% = $-0.8$ dB) of 19% (compared to the 18% overall). In the current test CBdB(50%) is also at -0.8 dB and there is a shift of 31% (compare to 10% overall). According to this analysis at equal sensitivity points the shift is larger with contralateral presentation of the precursor. These results show that the entire effect is central. This appears to be evidence for a central extrinsic compensation mechanism causing compensation with music-with-vocals. According to this measure there does not appear to be a role for a monaural effect. This result appears to suggest that compensation may have been suppressed when the precursor was presented to the same ear as the standard and comparisons. A reason for this could be that the enhancement effect was acting in the opposite direction to the spectral compensation effect. As crossaural presentation prevents the enhancement effect such a conflict would no longer occur.

Although equal sensitivity results show no role for a monaural effect in the shift with music, the overall measure shows a reduction in shift. This may be partly due to the fact that this measure includes continuum points for which ceiling effects occur. However, ceiling effects do not appear to be much greater in the crossaural test compared to the binaural tests so this might not fully explain this. It appears that there is a genuine reduction in the effect that may not be simply explained by increased ceiling effects. Therefore, the analysis of equal sensitivity CBdB points does not appear to give a complete picture of the effect. Analysis at other sensitivity points could also be undertaken.

It can be concluded that the effect with music-with-vocals is largely central and this is evidence of the spectral compensation effect occurring with this programme item. There is evidence for a reduction in effect with crossaural presentation of the precursors and therefore there may be a role for a monaural effect in compensation with music-with-vocals. Further tests should be undertaken to confirm this at a range of equal
sensitivity points. However, it is not necessary to perform these tests as part of this thesis. It can be concluded from the current data that a central effect occurs and there may be evidence for a monaural effect alongside this but this is not yet conclusive.

A conclusion of a largely central effect with evidence for a monaural component mimics that of Watkins (1991), whereby with a speech precursor a larger central effect was found with a smaller monaural component. The discovery of a central extrinsic compensation effect is strong evidence of the spectral compensation effect, as no other extrinsic central compensation effects, that can explain this shift, were uncovered by the literature review (but see Chapter 7 for a further discussion of this). As in Watkins’ (1991) study, the monaural component may in fact be caused by a central, direction sensitive process. Further research could be conducted to establish whether any of the monaural effect seen in the current test is robust to replication and whether or not it is a direction sensitive central effect or a genuinely monaural effect. If a direction sensitive effect is seen then this is further evidence for the spectral compensation effect.

**Music-with-vocals conclusion**

Compensation with music-with-vocals is shown to be partly central, as the shift remains with crossaural presentation of the precursor and comparison sounds. This result shows that an extrinsic central compensation mechanism must produce the effect. The only such mechanism revealed by the literature review was the spectral compensation effect. Therefore, the current results are evidence that the spectral compensation effect works with a non-speech musical programme item and is a general auditory compensation mechanism. However, the possibility of more speculative mechanisms explaining this effect, which were not researched as part of the literature review, is discussed in Chapter 7. These mechanisms are a central short time course effect and timbral memory. This result shows that the spectral compensation effect may contribute to real-world compensation and the time-gap sensitive component of this when listening to non-speech stimuli as well as speech stimuli.

There may be some contribution from a monaural effect to the shifts with the music-with-vocals programme item, as there is evidence for a reduction in effect with
crossaural presentation of the precursor compared to binaural precursor presentation, according to the overall measure. However, this was not shown when the shift was measured at equal sensitivity points using the CBdB(50%) measure. The overall measure provides the most comprehensive picture of this shift but it may be affected by ceiling effects, which may explain the reduction in shift seen here. Future work should aim to determine whether any monaural contribution is evident using the overall measure and measures at equal sensitivity points, when ceiling effects are prevented. The evidence for the monaural contribution to the shift is mixed but it is concluded that there is sufficient evidence of a monaural effect to conclude that this is likely to occur when listening to music-with-vocals. Ideally, further tests would aim to confirm this contribution but these are not conducted as part of this thesis. The monaural effect may show that the enhancement effect occurs with music-with-vocals. However, the result may also indicate a central direction sensitive process is behind the shifts seen. Direction sensitivity tests would also be necessary to confirm the existence of enhancement with this music item.

It was also noted that crossaural presentation of a precursor with monaural listening to the standard and test sound does not reduce the likelihood of achieving at least 70% accuracy responding at continuum endpoints in future tests. Therefore, it may be possible to compare compensation at at least this level of high sensitivity across the current and future experiments. Some bias was seen in the perception of the test continua however. Attempts should be made to eliminate this in future testing.

6.7.5 Results—Noise

The perception of HF content for the noise continuum, with +3 dB (‘More HF’) precursors and -3 dB (‘Less HF’) precursors is shown in Figure 6.14. This proportion of ‘More’ HF compared to the standard responses (y axis) is plotted against the continuum points (x axis), for each precursor condition +3 dB (blue) and -3 dB (red). The green function shows the average between the two precursors and represents listening with a neutral precursor. The expected monotonic trend of fewer ‘More’ responses at the -3 dB continuum endpoint to many ‘More’ responses at the +3 dB endpoint occurred. Nearly full psychometric functions were seen with highly accurate responses at continuum
endpoints for both precursor conditions.

![Figure 6.14](image.png)

**Figure 6.14**: The proportion of ‘More’ HF than standard responses at each S/C interval for the noise stimuli with crossaural precursor presentation. Functions are plotted for ‘More’ HF precursors (blue), ‘Less’ HF precursors (red) and the mean of the two (green).

**Baseline sensitivity and bias**

The data of primary interest to this study are the shifts caused by precursors with different filtering. This is measured by determining the extent to which the continuum shifts in the opposite direction of the filtering to the precursors (the shift of the blue and red functions), and by examining CB measures (see Section 6.7.5). However, the extent to which sensitivity is reduced or bias occurs due to the act of inserting a precursor (regardless of precursor filtering) should also be assessed.

**Sensitivity**

It can be seen that in a test where the standard and comparison sounds are compared monaurally and there is a contralaterally presented precursor accuracy remains high at continuum endpoints. Accuracy is 91% across both continuum points. As expected, sensitivity was reduced somewhat compared to monaural listening without a precursor (100%), showing that the task of comparing stimuli monaurally with a contralateral precursor is more difficult than listening monaurally without a precursor (significance
has not been tested due to low statistical power). Importantly, accuracy in this test has not fallen below 70%, so it is possible to make a comparison of compensation in this test and others at at least 70% sensitivity if desired.

Sensitivity appears to reduce by a larger amount with this item than with music-with-vocals and when monaural listening is done with a contralateral precursor compared to when monaural listening occurs without a precursor (9% reduction for noise and a 5.5% reduction for music-with-vocals). Contrary to expectation this shows an interaction between the programme item and the crossaural test conditions. However, the significance of the difference in reduction of accuracy with crossaural listening between music-with-vocals and noise is not tested due to low power so no conclusions are drawn.

**Pilot**

A pilot test was not previously conducted to determine whether the ±3 dB continuum is sufficient to measure sensitivity to at least 70% accuracy with noise when compared monaurally with an contralateral precursor in future tests. The same analysis as presented in Pilot 2 with monaural listening without a precursor is conducted for monaural listening with a contralateral precursor.

It is possible to pool both endpoints to give an endpoint accuracy of 91% with 95% CI at 72 - 97%. Sampling variation may have resulted in a higher accuracy in this test than the population mean accuracy. Given the 91% accuracy reported in the current test a non-inferiority test shows that with 95% confidence the true population value is not below 76% ($\Delta = 69\%$, $1 - 2\alpha \ CI (N = 28) = 76 - 96\%$) (Adjusted Wald). Given a true population value of not less than 76% it can be concluded that future tests with this continuum under the same conditions may yield values as low as 61% with 95% in tests with 35 participants ($\Delta = 69\%$, $1 - 2\alpha \ CI (N = 35) = 61 - 85\%$). Values below 70% may occur when a noise continuum is tested with monaural comparisons and contralateral precursors under these conditions in future tests. This is a conservative estimate. It is more likely that the true population value and sampled values will be higher, but based on the current data it might be desirable to extend the continuum when measuring noise in crossaural tests. However, it is possible that this estimate is
due to the low statistical power of this non-inferiority test. Repeating this pilot with more data points may give greater confidence that over 70% accuracy will be achieved in future tests. If after collecting more data the non-inferiority test still shows that sampled values below 70% may occur with more than 5% likelihood then the continuum may need to be expanded for this type of test.

Based on the information obtained it cannot currently be concluded that at least 70% will be reached in future tests. A larger reduction in sensitivity was seen in this test compared to music-with-vocals and this may indicate that the crossaural presentation of the precursor is particularly detrimental to sensitivity with the noise programme item and accuracy in this scenario may be genuinely reduced. However, regardless of the potential for this continuum to not yield the required sensitivity in future tests, high sensitivity (91%) is shown in the current test. Therefore, it is appropriate to analyse compensation effects. A comparison of this compensation at least 70% with other tests is possible.

**Bias**

There is some bias towards hearing the stimuli as containing more HF content with this continuum. This shift is described by CB measurements presented in Table 6.15. At CB% (0 dB) listeners responded ‘More’ 66% of the time. This is a shift of 16% from the expected 50% point. A two proportion z-test showed this not to be significantly different from expected ($z = 1.2, p < .225$). This is largely due to lower statistical power at 0 dB. The bias across all S/C intervals (Overall bias) was calculated by the proportion of ‘More’ responses across the whole test. This was significantly different to the expected 50% ($58\%, z = 2.2, p < .030$). The bias is most apparent in the asymmetric JND measures. The ‘Less HF than the standard’ JND is -1.6 dB but the ‘More HF than the standard’ JND is 0.2 dB, showing bias towards hearing the stimuli as containing more HF than the standard. It is not clear what causes this. There are even fewer reasons for bias to be experienced when listening to the noise item as the spectrum is uniform throughout the excerpts and onset sensitivity cannot occur - as may explain bias seen with the music items. It is unlikely that the bias is due to the average +1 dB shifting of the stimuli across the test due to roving, as there is no natural spectrum for the noise item. The bias may show a tendency for listeners to respond ‘More’ for task
reasons, such as choosing ‘More’ where they do not know the correct answer. However, no such tendency was reported by listeners. The bias is in the opposite direction to that shown with binaural comparisons (see Sections 6.5.6 and 6.6.12). It may be that the separation in time between the standard and comparison sounds that results from presenting a precursor to the other ear changes the bias present with this programme item compared to listening with an adjacent sound. In future crossaural tests with the noise item attempts should be made to reduce this bias.
Table 6.15: Bias measures for the HF continuum composed of the Noise with the LTAS of music-with-vocals programme item with monaural comparisons and a precursor presented contralaterally

<table>
<thead>
<tr>
<th>Measure</th>
<th>Expected</th>
<th>Result</th>
<th>Bias?</th>
<th>Significance of difference (2 tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CB%(Overall) Proportion of ‘More’ responses across the whole continuum</td>
<td>50%</td>
<td>58%</td>
<td>Bias towards more ‘More’ responses</td>
<td>( Z = 2.2, \ p &lt; .030 )</td>
</tr>
<tr>
<td>CB%(0 dB) Proportion of ‘More’ 0 dB comparison sound</td>
<td>50%</td>
<td>66%</td>
<td>Bias toward more ‘More’ responses</td>
<td>( Z = 1.2, \ p = .225 )</td>
</tr>
<tr>
<td>CBdB(50%) The location of 50% ‘More’ HF responding</td>
<td>0 dB</td>
<td>-0.4 dB</td>
<td>Bias toward more ‘More’ responses</td>
<td>-</td>
</tr>
<tr>
<td>CBdB(25%) location of 25% ‘More’ HF responding (the ‘Less HF than standard’ JND)</td>
<td>Symmetrical with CBdB(75%)</td>
<td>-1.6 dB</td>
<td>Bias towards more ‘More’ responses</td>
<td>-</td>
</tr>
<tr>
<td>CBdB(75%) The location of 75% ‘More’ HF responding (the ‘More HF than standard’ JND)</td>
<td>Symmetrical with CBdB(25%)</td>
<td>0.2 dB</td>
<td>Bias towards more ‘More’ responses</td>
<td>-</td>
</tr>
</tbody>
</table>
Shifts caused by precursors—0 dB continuum point

Figure 6.14 shows that precursors do not cause consistent shifts at most continuum points, the shift is frequently non-existent or in the opposite direction to that expected if compensation occurs. The shift at the continuum midpoint is large and in the opposite direction to that expected to be caused by compensation. Therefore there is no evidence of compensation according to this measure.

Shifts caused by precursors—Overall measure

A measure of the shift across all S/C intervals, is also obtained. The size of the shift overall was 0%. There is no significant difference between the conditions. This result is expected if the shift with noise is entirely monaural.

Shifts caused by precursors—Other continuum points

Figure 6.15 shows the shift at each continuum point with the 95% confidence interval for the shift. The results show no consistent shift. Many of the confidence intervals are centred around zero. This is indicative of no significant effect, as is shown with the overall measure.

Noise - reduction in effect

A reduction in effect with noise when the precursor is presented to the contralateral ear compared to the ipsilateral ear in Experiment 4 is evident. The CB(Overall) measure in Experiment 4 reported a 26% shift. The current experiment reports a 0% shift. This reduction is highly significant ($z = 10.5$, $p < .001$). There is no remaining shift with crossaural presentation. If compared at equal sensitivity points, the CBdB(50%) point for the binaural precursors is at +0.4 dB and the shift at this point is 53%. The CBdB(50%) point for the current test is at 0.4 dB where the shift is -4%, so the reduction is even greater when a comparison is made at equal sensitivity. The reduction in shift is evident at all continuum points and all equal sensitivity points.
Figure 6.15: The difference in the proportion of ‘More’ HF than standard responses at each S/C interval between ‘More’ and ‘Less’ HF precursors, for the Noise stimuli with crossaural precursor presentation. 95% confidence intervals for the difference are plotted.

It cannot be confirmed that the shift is exactly zero. The whole population would need to be measured to reach this conclusion; however, a two-sided test of equivalence can be run (Walker and Nowacki 2010). Based on the current data it can be concluded that the true population difference is not more than 7%. The current results and their marked divergence from data in the other experiments is a strong indication that there is no compensation with noise when listening crossaurally. This suggests an ear sensitive, direction sensitive or peripheral mechanism is the sole cause of the shift with noise and this is strongly suggestive of the enhancement effect occurring with noise.

**Noise Summary and Discussion**

The lack of a consistent shift suggests that the mechanism behind compensation with noise is monaural. It can be concluded that the data presented in the crossaural test with noise shows a marked difference in trend compared to other tests in this chapter. The results are strongly suggestive of an only monaural effect with noise. This is evidence for an effect caused by a peripheral mechanism and the enhancement effect. This appears to be the first instance of a test of the enhancement effect in a completely non-speech situation using broadband filtering that is similar to that caused by a
transmission channel such as a loudspeaker and room, rather than simple narrowband sounds that are used in masking tests. More data could be collected to support this conclusion. Data from the whole population would be necessary to show that the shift with crossaural presentation of precursors is zero but a test of equivalence where the delta is set at a point that is effectively zero should be run. This delta should be based on a shift that is biologically meaningless. However, even small shifts would probably have a role in perception (e.g. they might provide small compensation that enhances spectrally sensitive speech sounds) so it is not clear how small a shift would have to be in order to be regarded as meaningless. Any consistent crossaural shift that is found with a large dataset no matter how small is indicative of something other than a peripheral process as peripheral processes are entirely monaural.

The possibility remains that a central direction sensitive process is behind the shift seen with noise. This is unlikely as previous studies by Watkins showed that a monaural effect with noise was not caused by a direction sensitive process. However, this should be confirmed for the current test. In Watkins’ test a speech test sound was used. It may be that any direction sensitive effect that might be present with noise did not occur in that study due to ungrouping between the precursor and the test sound because they were different. In the current study the precursor and test sound were both the same programme item so this could not occur. However, ungrouping between a non-speech programme item and speech test sound did not prevent the spectral compensation effect in Holt’s (2005, 2006) studies. Further, if a direction sensitive effect is not shown when speech is present and where a special mechanism may occur, it is even less likely that it will occur in an all noise test where the mechanism is more likely to be a general auditory process and peripheral. It is concluded that it is possible but unlikely that the monaural effect seen with noise is a result of a direction sensitive central effect.

The crossaural results with noise mimic results with noise seen in Watkins’ (1991) tests. The results provide further evidence that a monaural compensation process that is a general auditory process occurs when listening to noise. This may contribute to channel compensation. However, if the effect is the enhancement effect, it is not clear to what extent this short time course effect would provide channel compensation.
Noise conclusion

The data presented in this crossaural test with noise shows a marked difference in trend compared to the other tests conducted in the chapter. When listening to noise there is no shift caused by precursors when the precursor stimulus is presented to a different ear to the test sound. This suggests that only a monaural mechanism contributes to compensation with noise. This result is evidence that the enhancement effect occurs in this test. However, it is possible but unlikely that a central direction sensitive process is behind this effect. The findings in this test support those of Watkins which suggest that static stimuli bring about only a monaural compensation process.

It was also noted that crossaural presentation of a precursor with monaural listening to the standard and test sound may reduce the likelihood of achieving at least 70% accuracy responding at continuum endpoints in future tests. This should be confirmed. If this remains the case then a wider test continuum may need to be used with noise under these test conditions to ensure that compensation can be measured at at least 70% accuracy across the current and future experiments. A small but significant bias was seen in the perception of the test continua. Attempts should be made to eliminate this in future testing.

6.7.6 Experiment 5: Summary and discussion

The main aim of Experiment 5 was to determine whether the shifts seen in Experiment 4 were a result of a monaural mechanism, crossaural mechanism or both, as this would suggest that either the enhancement effect, the spectral compensation effect, or both cause compensation with non-speech sounds.

As expected there was no spectral compensation effect with noise as shifts with noise stimuli did not occur crossaurally. This finding is consistent with that of Summerfield and Watkin’s crossaural tests with noise and static harmonic complexes. This result is a strong indication that a peripheral neural adaptation mechanism and the enhancement effect cause compensation with noise. It remains possible that a central direction sensitive process is behind this effect, but Watkins has shown that the monaural effect
with noise is not direction sensitive. Ideally, a direction sensitivity test should be run to confirm that compensation with noise is caused by the enhancement effect, however, this is not conducted as part of this thesis as previous research described in Chapter 5 suggests that a finding of only a non-crossaural effect with noise is expected and so the results here are strongly indicative of the monaural auditory enhancement effect causing compensation.

If the enhancement effect is the only effect that causes the shifts with noise, then it may be the only extrinsic mechanism that causes real-world compensation when listening to noise, and the only time-gap sensitive mechanism at work when listening to noise. This may mean that in a real-world test where noise is the programme item, compensation may occur but via enhancement only, or at least a time-gap sensitive effect would occur but via enhancement only. The spectral compensation effect does not appear to contribute to real-world compensation with noise but other non-time-gap sensitive mechanisms may also cause compensation in real-world experiments with noise.

When listening to music, compensation appears to be largely the result of a central effect. Therefore, the spectral compensation effect is confirmed in an all music test. With overall measures of compensation across the test sound continuum there is a significant reduction in effect with contralateral compared to binaural precursor presentation, suggesting that a monaural effect also contributes to compensation with music. It is possible that any reduced effect seen is due to ceiling effects that can occur with overall measures. However, ceiling effects do not appear to be much more prominent with the crossaural test compared to the binaural test so there appears to be a genuine reduction in effect with crossaurally presented stimuli when measured across the whole continuum. The reduction in effect does not necessarily indicate that the enhancement effect contributes to compensation with music-with-vocals. The reduction may be due to a central and direction sensitive process. Evidence of a direction sensitive process would be further evidence of the spectral compensation effect causing compensation in a music test. However, it is expected that the enhancement effect would contribute to compensation with music to some extent, as this was seen with speech in Watkins’ (1991) test. If enhancement has been shown to contribute to shifts in a test where a special mechanism is more likely to dominate, it may be even more likely to contribute in a non-speech test where general auditory processes are
However, it should be noted that when equal sensitivity points at 50% sensitivity were compared there was no evidence for a monaural effect. In fact compensation was increased rather than reduced with crossaural listening. However, the shift at this point does not appear to be representative of that across the whole continuum and at other equal sensitivity points there is a reduction in effect. Therefore, it is concluded that the findings for a monaural effect with music are mixed but the most persuasive measure, the overall measure, is suggestive of this, so it is likely that this occurs. If future experiments show that there is no reduction in shift with crossaural listening, and in fact the shift is larger, then this might be due to a cancelling out of the enhancement effect and spectral compensation effect because the prior phonemes’ worth of music might have a different LTAS to the whole segment. This was considered as an explanation of the reduction of the shift with music compared to noise in Experiment 4. Further work would be needed to confirm whether the effects are greater or reduced with crossaural listening compared to binaural listening before the issue of cancelling out is examined further. This work will not be done as part of this thesis as the overall measure provides good enough evidence of a reduction in effect.

It was noted at the beginning of this chapter that if these effects occur with non-speech and then Step 1 Questions A-C (set out at the beginning of this chapter and in Chapter 4) can automatically be answered for these mechanisms with non-speech listening (see also Section 5.4.2). It is therefore now possible to conclude that the spectral compensation mechanism and the enhancement mechanism are both potential mechanisms of the time-gap sensitive component of compensation when listening to speech and non-speech sounds. These mechanisms may therefore explain the time-gap sensitive component of compensation seen in real-world listening to speech and the compensation in Olive et al.’s (1995) real-world test with music. Research Question 1b has been answered. Future work can now proceed to further confirming the role of these effects in real-world listening using the real-world paradigm (Step 2 of the research process set out in Chapter 4 can be conducted). Future work can also examine the underlying mechanisms of the non-time-gap sensitive component of compensation using literature reviews (like that described in Chapter 5, fulfilling Step 1b(i) of the research process), further psychoacoustic laboratory tests (like those described in this chapter,
fulfilling Step 1b(ii)), and real-world tests like those described in Chapter 3 (fulfilling Step 2b). Future work is discussed in the next chapter.

In preparation for some of the future work discussed in the next chapter, pilot tests were run to confirm the suitability of the stimuli range used to measure compensation in the current test and in future tests. In Experiment 5, further analysis was undertaken to measure the suitability of a $\pm 3\text{dB}$ continuum for measuring compensation at equal high and low sensitivity points across tests. It was shown that any test with monaural listening with a contralateral precursor may result in accuracy below 70% at continuum endpoints in future tests with noise. This result may be due to the low number of data points leading to a conservative estimate but based on the current data a $\pm 3\text{dB}$ continuum is not sufficient to show that a reduction in accuracy below 70% will not occur with 95% confidence in future crossaural tests with noise. Further analysis should be done to ensure that there is not something about listening to noise with a contralateral precursor that means that sensitivity is likely to be low in future tests as this may prevent comparisons of compensation effects at an equal high sensitivity point between tests. As noted in Experiment 4, it will be possible to measure compensation at low sensitivity with this type of continuum so measuring compensation at low sensitivity is not an issue. Bias was present for both music-with-vocals and noise. Attempts should be made to remove this before future work, as bias suggests unexpected perceptions of the stimuli which may influence test results and may cause ceiling effects, which prevent accurate measurement of compensation.

### 6.7.7 Experiment 5: Conclusion

Two mechanisms of compensation have been identified with non-speech: the enhancement effect and the spectral compensation effect. It is confirmed that these are time-gap sensitive and Questions A-C have been answered for these mechanisms. It can be concluded that these are potential mechanisms of the time-gap sensitive compensation for spectral colouration caused by loudspeakers and rooms when listening to speech and the music seen in Olive et al’s original experiment. More specifically, it can be concluded that when listening to noise through loudspeakers and rooms the spectral compensation effect is not a potential mechanism of compensation but when
listening to music it is. Research Question 1b has been answered and Step 1a of the research process is complete. Future work can be conducted to continue this research process and further examine real-world compensation for spectral colouration caused by loudspeakers and rooms.

6.8 Chapter summary and discussion

This chapter described Experiments 4, 5 and the associated pilot experiments. The main aim of Experiment 4 was to determine whether shifts indicative of compensation occur with non-speech sounds. If shifts in the perceived spectrum of non-speech test sounds occur, and these shifts are in a contrastive direction to the filtering applied to non-speech precursors, then this is suggestive of an extrinsic mechanism (either the enhancement effect or the spectral compensation effect) providing compensation for channel colouration in non-speech perception. Experiment 5 aimed to test the mechanisms behind the compensation seen in experiment 4 and determine whether the enhancement effect, the spectral compensation effect, or both are behind these extrinsic shifts. No other compensation mechanisms were tested as it was noted in Section 5.4 that only these extrinsic processes are likely to explain the time-gap sensitive component of compensation, so only these needed to be tested with non-speech.

Before the main tests were conducted pilot tests were run to validate the range of test sounds to be used in Experiments 4 and 5 and in future tests to ensure they elicited the expected perceptions and covered an appropriate range to be able to measure compensation. Specifically, the aims of these pilot experiments were to ensure that the test sound continuum produced a range of sounds that varied in a monotonic manner in brightness, so that shifts in brightness with compensation could be measured. Listeners did not perceive the sounds to range in brightness but the brightness continuum was heard to vary in HF. Therefore, shifts in perceived HF content were used to measure compensation in the tests described in this chapter. A continuum of sounds ranging in HF content was used to examine how the magnitude of compensation varies depending on the degree of difference between the standard and comparison sounds (i.e. to measure compensation at different underlying sensitivity to timbral
6.8 CHAPTER SUMMARY AND DISCUSSION

differences). It was expected that compensation would be larger at low sensitivity and small at high sensitivity. It was acknowledged that in tests of this type it would always be possible to measure compensation at low sensitivity, as chance performance is expected somewhere along the continuum where the continuum contains the same sound and the standard sound. Therefore the analysis focused on the ability to achieve a measure of compensation at high sensitivity points. The pilot tests ensured that compensation could be measured at at least 70% (with greater than 95% confidence) accurate responding with each of the programme items, in a variety of test conditions. The continuum range selected (± 3dB) appears to be appropriate for measuring compensation at relatively high sensitivity with difficult to discriminate programme items like music and sufficient to measure compensation at low and medium sensitivity with easy to discriminate programme items like noise. This range was concluded to be appropriate for use in Experiments 4 and 5 but also deemed appropriate to be used across a variety of tests with different programme items and test conditions that might be run in the future. It should therefore not be necessary to expand the range in future testing. However a further test pilot test with noise may be necessary to confirm the suitability of this continuum range. Establishing a fixed range to be used in future tests helps to prevent range effects varying between tests which are known to affect perceptual judgements (Poulton 1989), and it allows for direct comparison of results from future tests at equal high sensitivity points. At least 70% sensitivity will be reached across tests and it is likely that in practice high sensitivity points of above 80% will be reached with the continuum range selected. In this case comparisons of compensation at this point can also be made. The pilot tests also aimed to check that bias would not affect the perception of the continuum to the degree that it might result in ceiling sensitivity preventing compensation being measured at a range of sensitivity points. Bias also distorts compensation as revealed in overall measures. Further, bias is undesirable because it represents effects on perception that are not elicited or controlled by the test. A small but significant bias was seen in almost all test conditions. It was recommended that attempts should made to eliminate this in future tests to allow for a more accurate measure of compensation. However, the bias was not large enough to prevent a valid assessment of compensation in the current tests.

The aim of the main experiments (Experiment 4 and Experiment 5) was to determine
whether timbral shifts indicative of compensation occur with non-speech sounds when using the traditional psychoacoustic laboratory paradigm, rather than the real-world paradigm. The reasons for using this paradigm were given in Chapter 4. This paradigm was used to first show that the extrinsic mechanisms of compensation revealed in the literature, which may explain the time-gap sensitive component of compensation (the enhancement effect and the spectral compensation effect), can occur with non-speech as well as speech. Only when these mechanisms are confirmed with both programme items can they be considered general auditory mechanisms of real-world compensation. Experiments 4 and 5 therefore aimed to fill gaps left in the literature review so that potential mechanisms of the time-gap sensitive component of real-world compensation when listening to non-speech sounds could be determined (i.e. they aimed to complete Step 1a(ii) of the research process set out in Chapter 4).

In Experiment 4 it was shown that compensation occurs with non-speech sounds via an extrinsic mechanism and therefore either the enhancement or the spectral compensation effect occurred with non-speech. As was discussed in Section 5.4, these mechanisms are compatible with explaining the time-gap sensitive component of compensation. The existence of compensation with non-speech items indicates that compensation does not rely on speech-specific mechanisms. This result fits with studies showing compensation with non-speech precursors or non-speech test sounds. The size of the shift varied depending on the programme item was but was 22% of the continuum on average. The shifts were similar in magnitude with all items but there was a significantly reduced shift with music compared to noise. There was no apparent benefit of vocals in the programme item suggesting either no special compensation with speech or that vocals are not treated as speech. The results do not preclude compensation being larger with speech. Regardless of the music-with-vocals being unlikely to be perceived as speech, it is still surprising that this item did not bring about a larger shift than noise. This item and the instrumental item contained spectral variation which, according to Watkins’ hypothesis, brings about an additional and larger spectral compensation effect compared to that that occurs with static stimuli like noise. The lack of a larger effect with music suggested that only the enhancement effect occurred with the non-speech programme items used in this test. However, it was noted that the effect sizes are not particularly informative as to the mechanisms of compensation. The crossaural nature
is the decisive factor in distinguishing mechanisms.

Experiment 5 involved crossaural tests to examine whether a monaural or central effect was behind the shifts seen in experiment 4. If a monaural and/or central effect is shown it may be concluded with some certainty that enhancement and/or spectral compensation cause compensation with non-speech. Experiment 5 shows that shifts with noise did not occur crossaurally. This finding is consistent with that of Summerfield and Watkins' crossaural noise and static harmonic-complex precursor tests and suggests that only enhancement causes compensation with noise precursors. Crossaural shifts occurred with music showing that the spectral compensation effect occurs with music. Therefore, it appears that a lack of spectral compensation with noise is probably not due to the non-speech nature of the stimulus but its lack of spectral variation. It is possible, but unlikely based on Watkins' 1991 study which showed no direction sensitive effect with noise, that a central direction sensitive process is behind the lack of crossaural effect with noise. It can be concluded that when listening in the real world any time-gap sensitive component with noise may be due to enhancement but not the spectral compensation mechanism. Additionally, any non-time-gap sensitive contribution to real-world compensation cannot be from the the spectral compensation effect. In the future it may be useful to conduct a real-world test with noise to examine the extent of compensation in this scenario. However, this is not conducted as part of this thesis.

With music both the spectral compensation effect and the enhancement effect appear to cause shifts indicative of compensation. The shift remains significant when stimuli are presented crossaurally, indicating the spectral compensation effect affects listening with music, but the shift it is significantly reduced compared to monaural presentation, indicating that the enhancement effect also acts with music. As suggested in the discussion of the results with noise, the monaural component should be further confirmed as it is possible, but unlikely, that it is caused by a direction sensitive central effect. Additionally, a comparison of compensation at the equal sensitivity points at 50% showed no monaural effect. However, overall measures which are more representative of listening with a range of sounds showed a reduction with crossaural listening, so it is concluded that it is likely that enhancement occurs alongside spectral compensation when listening to music.
Future tests may show that there are differences in effect size between non-speech and speech in terms of both enhancement and spectral compensation. Further, a speech specific mechanism may still occur to engage extra compensation compared to that seen with non-speech. Any differences in effect sizes between speech and non-speech may due to a central cognitive process that occurs when the listener is aware of the sounds being speech rather than acoustical differences between speech and music. However, acoustical differences between speech and music do exist, and these could cause differences in compensation effect sizes. It would be desirable to make direct comparisons between programme items in future tests. It is not possible to make a comparison with Watkins et al’s shifts with speech as the stimuli are too different to those used in this test. Future tests should aim to test speech alongside, music (with and without vocals) and noise.

6.8.1 Chapter questions

The chapter questions are now addressed:

**Question 6.1** asked: ‘to what extent does the work in this chapter show that any of the specific compensation mechanisms cause compensation in non-speech perception?’

Experiment 4 showed shifts indicative of an extrinsic compensation mechanism in entirely non-speech tests. The only established mechanisms that can cause such compensatory shifts are the enhancement and spectral compensation effects. It was therefore shown that either enhancement and/or spectral compensation occur with non-speech (music and noise).

**Question 6.2** asked: ‘exactly which of the compensation mechanisms result in this compensation with non-speech?’

Experiment 5 showed that, as expected, noise appears to only elicit a monaural compensation effect. This is strongly suggestive of the enhancement effect occurring with noise and no contribution from a central mechanism. However, although unlikely it remains possible that the effect was caused by a central direction sensitive process. With music a central component to the shifts was seen. This appears to show an
central extrinsic compensation mechanism working in a non-speech context. The component of the shift that is definitely central is likely to be caused by the spectral compensation effect. A monaural effect, indicative of the enhancement effect also appeared to contribute to compensation with music. Further testing which limits the possibility of ceiling effects and tests for direction sensitivity could be undertaken to confirm the existence of this monaural effect but the balance of evidence suggests that the enhancement plays a role in compensation with music.

**Question 6.3** asked: ‘are any of these mechanisms potential mechanisms of compensation for loudspeakers and rooms in real-world listening to non-speech?’

The results from experiment 5 confirm the existence of the enhancement effect and the spectral compensation effect when listening to non-speech. Further, in Chapter 5 and in the summary and discussion section for experiment 5 (Section 6.7.6) it was noted that upon finding these effects it could be assumed that the criteria for concluding that these effects are potential mechanisms of the time-gap sensitive component of real-world compensation for loudspeakers and room in real-world listening (time gap sensitivity and Step 1 Questions A-C being answered in the affirmative) would be automatically met. This is because these mechanisms have already been shown to be unique, and have a nature and time course that is appropriate to explain real-world compensation and the time-gap sensitive component of this. There is nothing about non-speech listening that is expected to change the features or time course of these mechanisms to the extent that this is not longer the case. It is therefore concluded that the enhancement effect and the spectral compensation effect are potential mechanisms of the time-gap sensitive component of compensation for spectral colouration caused by loudspeakers and rooms in real-world listening.

However, it is acknowledged that real-world tests could be conducted to confirm the role of these mechanisms in such compensation. Further, it appears that a central extrinsic mechanism such as spectral compensation does not have a role in compensation with noise in laboratory tests (i.e. it is not a potential mechanism of compensation for loudspeakers and rooms when listening to noise). Therefore, it will not have a role in real-world compensation with noise.
Both the spectral compensation effect and enhancement effect have the potential to play a role in real-world compensation with music. It appears that the majority of real-world compensation with music may be provided by the spectral compensation effect, or at least the majority of the time-gap sensitive component of real-world compensation with music may be explained by this. This is because a) these appear to be the only two known compensation mechanisms that appear to be able to explain the time-gap sensitive element of real-world compensation and b) the spectral compensation effect is longer lasting and thought to be larger than enhancement. However, it does still remain possible that other compensation mechanisms exist that could explain this component of compensation, but these were not uncovered by the literature review and so were not investigated here. In section 7.2.3 the potential role of timbral memory, which has been dismissed as a compensation mechanism in this thesis, is discussed further.

As was shown in Experiment 3 non-time-gap sensitive mechanisms are likely to have a large role in real-world compensation with both music and noise. Further work would be necessary to examine the mechanisms behind the non-time-gap sensitive component of compensation. It is possible that the spectral compensation effect also contributes to difference between direct and indirect comparisons that is not to do with time gap (i.e the non-time-gap sensitive component of real-world compensation).

## 6.9 Chapter conclusion

Experiments 4 and 5 aimed to confirm that extrinsic mechanisms that may contribute to the time-gap sensitive portion of compensation—the enhancement effect and the spectral compensation effect—do not arise as part of a speech mode of perception but occur with non-speech as well as speech. These experiments were necessary to confirm potential mechanisms of the time-gap sensitive component of compensation for loudspeakers and rooms when listening to non-speech, including the music used in Olive et al’s original test.

Both effects were shown to occur in laboratory tests with the same music used in Olive et al’s original experiments, but the spectral compensation effect did not occur with noise. The criteria for determining a potential mechanism of the time-gap sensitive
component of real-world compensation (Step 1 Questions A-C being answered in the affirmative and time-gap sensitivity) were immediately met upon finding these mechanisms with non-speech. It could be concluded that these criteria were met for these mechanisms because they had been met when listening to speech and the features and time course of these mechanisms are not expected to change with non-speech, or at least not to the point that the mechanisms are no longer: unique (a distinct compensation mechanism—see Question A); compensatory in nature (causing the shifts in perceived spectrum towards the perceptual midpoint during listening, and enhancing differences between adjacent stimuli where channels change, as seen with indirect comparisons in real-world tests—see Question B); extrinsic (not immediately acting but providing compensation with a longer time course of listening—see Question C); and time-gap sensitive (enhancing timbral differences between adjacent sounds when there is little or no time between listening but not enhancing (hence compensating) with time between sounds). Therefore, Research Question 1b has been answered:

‘By what means can compensation affect the perception of loudspeaker and rooms?’

Both the spectral compensation effect and the enhancement effect are potential mechanisms of the time-gap sensitive component of compensation for loudspeakers and rooms when listening to music and speech but only the enhancement effect is a potential mechanism of this compensation with noise.

Further work following the work in this thesis may now be conducted to confirm the actual contribution of the potential compensation mechanisms to real-world compensation using the real-world paradigm (i.e. Step 2a of the research process set out in Chapter 3 can be conducted) and the non-time-gap sensitive component of real-world compensation can be investigate (Steps 1 and 2 b can be conducted). This work is discussed further in Chapter 7. Work has already been done to ensure that the results of any future tests, conducted using the laboratory paradigm, can be compared with the results seen in the tests in these chapters. It has been ensured that the range of sounds used in these tests will be appropriate to show compensation at a range of sensitivities. Compensation effects can be compared at equal high and low sensitivity points across tests, allowing for a more valid comparison of compensation effect sizes across different conditions.
Chapter 7

Discussion and future work

This thesis aimed to determine whether compensation for transmission channels, specifically loudspeakers and rooms, occurs in real-world listening. It also aimed to determine the potential mechanisms behind this compensation. The current chapter will summarise the work conducted and discuss future research that might be carried out to continue the research process begun here. This chapter is divided into 3 sections:

Section 7.1 will summarise the work presented in this thesis and the answers to the chapter questions. Section 7.2.1 will make suggestions regarding future work to further determine answers to the research questions set out here. Section 7.3 concludes this thesis by laying out the main findings and showing how the research questions have been answered. A bullet point list of the contributions of this work to the research field is given.

7.1 Summary of work so far

In the introduction to Chapter 1 the research problem was described. The broad aim of the research in this thesis—to investigate compensation for spectral colouration caused by loudspeakers and rooms—was set out. It was noted that these specific channels would be examined because they are representative of general transmission channels and can shed light on how the hearing system functions despite distortions caused by
any channel and the natural environment. It was also observed that compensation for these channels is of interest to those in the audio industry who want to understand the extent to which channels such as rooms and loudspeakers can be heard and how to produce audio products/machine listeners that can adapt to the environment like a human listener. Loudspeakers and rooms are often tested in laboratory studies where they are shown to have a large effect on the sounds passing through them. However, compensation may take time to occur. In the real world, amongst other things, there are differences in the time course of listening compared to laboratory studies (people listen to a single channel for longer and hear different channels with time between them, unlike in laboratory listening). Compensation for the effects of loudspeakers and rooms may be evident due to compensation that occurs over this longer time course of listening or due to other aspects of real-world listening. It was noted that more work is needed to examine compensation in real-world listening. The work in this thesis aimed to determine whether compensation for loudspeakers and rooms occurs during real-world listening and the extent of this compensation. It also aimed to determine the potential mechanisms by which this occurs. It was noted that specific compensation mechanisms appear to exist that reduce spectral colouration caused by the vocal tract (VT) and phonetic context in speech perception. The VT and phonetic context colour the sound in a similar way to other transmission channels so these mechanisms might also provide compensation for loudspeakers and rooms when listening to speech or non-speech. However, the extent to which these mechanisms cause real-world compensation for loudspeakers and rooms is not clear. They may be speech specific or occur only to compensate for these specific distortions. Further, research into these mechanisms has only been conducted over short time periods so they may not occur at all or be an unimportant means of compensation during the longer time periods of real-world listening.

Specific research questions, based on the aim of establishing the extent of real-world compensation for loudspeakers and rooms (Research Question 1a) and its potential mechanisms (Research Question 1b) were set out. Question 1c was also posed, which asked whether any of the potential mechanisms uncovered by answering Research Question 1b actually contribute to real-world compensation. Answering this question is important to confirming the causes real-world compensation but it was declared
that the work in this thesis will only contribute to establishing potential mechanisms of compensation. After the research questions were introduced chapter questions were described which structured the information presented in the remainder of the chapter towards further defining the research problem and providing answers to Research Questions 1a and 1b. The work in the remainder of the chapter can therefore be summarised via the answers to the chapter questions.

Question 1.1 asked how timbre is used in sound recognition in speech and non-speech perception. Timbre is the perception of a sound’s character and a key cue to identifying and recognising a sound. In speech perception timbre allows us to recognise different phonemes. The timbre of a complex sound is largely determined by its spectral content. Speech timbre is defined by its spectral content and in particular by spectral resonance peaks known as formants. The size and the shape of the instrument body determines the timbre of non-speech sounds and the vocal tract size and shape determines the spectrum of speech sounds and formants.

Question 1.2 asked about the effect of the transmission channel on timbre and sound recognition. It was noted that the environment in which a sound is heard or the transmission channel (e.g. microphone, loudspeaker, listening room) through which a sound passes can distort its spectrum. This distortion occurs due to energy being amplified and attenuated by different degrees in different frequency regions due to channel resonances and reflections. Transmission channels such as loudspeakers and rooms colour the sounds passing through them in a similar way to the vocal tract or other instrument body. This results in distortion to the timbre of a sound source which can prevent auditory object recognition.

Question 1.3 asked: ‘what is timbral constancy and how does this occur to reduce the effect of the channel to allow for more accurate auditory object identification?’. It has been observed that colouration caused by transmission channels does not impede listeners’ ability to recognise sounds. The effects of the channel spectrum appear to be perceptually reduced relative to sounds passing through it. This timbral constancy is thought to be the result of one, or a number of, perceptual processes, which may include mechanisms that compensate for the transmission channel spectrum. Mechanisms that provide timbral constancy may originate peripherally or centrally in the hearing system
and they may be physiological or cognitive. It was noted that mechanisms exist that appear to compensate for colouration by the VT in certain circumstances and these may be the same as, or similar to, mechanisms that compensate for other channels. It was observed that timbral constancy and compensation is unlikely to completely remove colouration by the channel and some perception of the channel is likely to be useful for identifying the channel and or the talker that produced a speech sound.

Research Questions 1a and 1b were set out but not answered in this chapter. However, upon answering the chapter questions it was concluded that compensation appears to exist and this provides timbral constancy in listening and there is a suggestion that there is something about real-world listening that brings about increased compensation compared to laboratory listening. A longer listening time course might be behind real-world compensation. It was also noted that timbre is processed at multiple levels in the hearing system so compensation mechanisms could act at a number of points along the stimulus-to-perception chain. It was concluded that work in the remainder of the thesis will aim to report previous research and conduct additional experiments to provide answers to research Questions 1a and 1b.

7.1.1 Chapter 2

An aim of Chapter 2 was to review evidence for real-world compensation for loudspeakers and rooms from audio engineering experiments and examine its extent. A number of these studies have shown that these channels have large perceptual effects on sounds passing through them. This suggests a lack of compensation. However, a few studies have shown that listeners are less sensitive to these effects in certain circumstances compared to others and compared to what is expected based on objective measures, suggesting compensation. The findings from this chapter are summarised in the answers to the chapter questions:

**Question 2.1** asked: ‘what is the evidence for timbral constancy and compensation for the channel in real-world listening to loudspeakers and rooms from prior experiments?’ There is evidence for compensation for the timbral effects of both loudspeakers and rooms when listening to music and other material reproduced through these
channels. Compensation may occur instantly via a binaural rejection mechanism but
the contribution of this appears small as loudspeaker and room effects are reported
to be large. Loudspeaker and room effects appear further diminished with a longer
listening time course compared to that common in laboratory listening (i.e. when the
channel is heard unchanging for longer and/or when hearing each channel with time
between it and other different channels). A longer listening time course may cause
extra compensation in real-world listening but extra compensation may also occur due
to other differences between laboratory and real-world listening.

Only a couple of studies (Olive et al. (1995) and Olive and Martens (2007)) have
specifically aimed to measure compensation in the real-world scenario, comparing
listening over a longer time course to an appropriate baseline: listening to the same
stimuli over a short time course, as in laboratory listening. Of these studies only one
(Olive et al. 1995) showed increased compensation for loudspeakers and rooms in the
real-world scenario compared to the laboratory scenario. Failure to find real-world
compensation in Olive and Martens’ (2007) study appears to have been due to task
factors. It was concluded that Olive et al.’s (1995) study should be replicated to
confirm real-world compensation for loudspeakers and rooms and measure its extent.
It was observed from the studies presented in this chapter that neither real-world or
the instant compensation that affects laboratory listening is complete.

**Question 2.2** asked: ‘what is the best way of measuring real-world compensation for
loudspeakers and rooms in an externally valid manner and how will compensation be
measured in this thesis?’. An aim of chapter 2 was to establish the most appropriate
test paradigm to use to confirm that real-world compensation occurs and test its
mechanisms. Olive et al.’s paradigm provides a means of measuring compensation
that occurs over longer and shorter listening time courses and in situations otherwise
more similar to real-world listening, (including measuring both the effect of longer
continuous listening and longer time gaps between listening, both of which may
contribute to compensation in the real world). This paradigm directly compares the
effect of loudspeakers and rooms in the laboratory listening scenario when measured
using a direct comparison method (short listening time course) to one that is more like
real-world listening, measured using an indirect comparison method (longer listening
time course). Thereby an appropriate baseline measure is provided. This paradigm
is preferable to examining perception against objective measures which would not measure real-world compensation over longer time courses and may not measure compensation at all. This paradigm also allows for simultaneous perception of two factors: the loudspeaker and the room. This is more similar to real-world listening in light of multiple simultaneous factors compared to laboratory listening that often involves listening in a more isolated way. It was concluded that this paradigm is suitable for further confirming real-world compensation and testing its mechanisms and will be used for the work in this thesis. However, the paradigm does not allow full control of listening time courses or allow measurement of very long or short compensation so future work may need to modify the paradigm to some extent.

The work in this chapter contributed to answering Research Question 1a. Compensation for the spectral colouration caused by loudspeakers and rooms appears to occur in situations similar to real-world listening that does not occur in laboratory listening. The most prominent difference between real-world listening and laboratory listening is the longer time course and this may be behind this compensation. This compensation must be confirmed as it was only revealed in one study. Real-world compensation does not appear to provide complete channel compensation. Further research is necessary to fully answer Research Question 1a and confirm real-world compensation for loudspeakers and rooms and measure its extent. The research in this chapter did not aim to directly answer Research Question 1b but some of the information presented contributed to determining the mechanisms of compensation. It was hypothesised by Olive et al. (1995) that real-world compensation is due to the time gap between listening with the indirect comparison method (i.e. in real-world listening). Specifically, memory may be behind this compensation. It was also noted by this author that indirect comparisons involved longer periods of continuous listening to an unchanging channel and a mechanism that is sensitive to longer listening periods might cause real-world compensation, such as a process of timbral fatigue. However, these hypothesised mechanisms of compensation are speculative. Specific compensation mechanisms may exist that can explain real-world compensation. Upon confirming real-world compensation, the mechanisms behind this should be investigated further.
7.1.2 Chapter 3

Only two experiments have been conducted that measure loudspeaker and room compensation using a real-world listening paradigm, which compares listening over the longer time courses, as in the real-world, and short listening periods, as seen in the laboratory (Olive and Martens 2007, Olive et al. 1995). Of these studies one showed compensation for loudspeakers and rooms (Olive et al. 1995). Three experiments were conducted to replicate and expand the research of Olive et al. (1995). Experiment 1 aimed to replicate the results of Olive et al. and show compensation in real-world listening. Specifically, like Olive et al.’s original experiment this experiment aimed to determine whether the effects of loudspeakers and rooms would be perceived to be reduced when making indirect comparisons of different channels (involving listening continually to the channel for longer time periods and time between comparisons of different channels), compared to direct comparisons (involving listening to the channel for short time periods and comparing the channels with no time gaps between them). This study showed some compensation but this did not reach the required level of statistical significance that is necessary to firmly conclude that compensation occurs, therefore Experiment 2 aimed to replicate Olive et al.’s (1995) findings with conditions more conducive to compensation. Real-world compensation was found so Experiment 3 aimed to determine the causes of the compensation seen in Experiments 1 and 2 and in Olive et al.’s (1995) original study. The answers to the chapter questions further summarise this work:

Question 3.1 asked whether compensation for loudspeakers and rooms seen in the real-world experiments described in chapter 1 could be elicited again with new stimuli. Experiment 1 showed evidence of real-world compensation for the loudspeaker factor but this was only significant at p<.10. Compensation for the room factor could not be measured because of floor effects but a trend towards compensation was seen. It was concluded that compensation should be tested again using a wider range of stimuli and conditions more conducive to compensation to confirm compensation for both loudspeakers and rooms. Experiment 2 aimed to replicate Olive et al.’s (1995) findings with conditions more conducive to compensation. To prevent floor effects, the perceived differences between the loudspeakers and rooms were enhanced by using
a speech programme item. Increased data points and a longer practice session also aimed to reduce error in measurements. To increase the size of the compensation a longer time gap between listening in the indirect comparison condition was used and the length of continuous listening was increased to be similar to the length used in Olive et al.’s work. There was a significant difference between loudspeakers and rooms when compared directly and this difference decreased when compared indirectly. Highly significant compensation was observed for both the loudspeaker and room factors. Experiment 3 also showed compensation. Overall, there is now evidence of compensation occurring with speech and non-speech programme items, both replicating Olive et al.’s (1995) results and increasing the validity of Olive et al.’s (1995) results to another programme item. An aim of the research in this thesis is to measure the extent of compensation. This work quantified compensation more precisely, using an ecologically valid measure—the preference rating scale—which measures timbral effects in terms of large, moderate and small effects as decided by the listener. The compensation seen for the room factor in Experiment 2 in conditions conducive to compensation was large according to rating scale values (29 preference points) and the compensation seen for the loudspeaker factor was moderate (10 preference points) according to rating scale values.

In order to answer Chapter Question 3.2: ‘Is this compensation an artefact of the experimental process or is it likely to be a genuine listening phenomenon?’, it was necessary to confirm that the effect seen in the previous studies is a genuine compensation effect and not an artefact of the experimental process. Specifically, it was necessary to confirm that results were not due to task factors and instructions causing a bias towards listeners ignoring differences in timbre across the test (i.e. differences in timbre between indirectly compared stimuli). To remove any bias specific instructions were given to the listeners to pay attention to changes due to the indirectly compared channels changing across the test and mark these in ratings. Listeners were either instructed to make ‘global ratings’ and mark any differences heard across experimental pages or were given no specific instructions to do this (i.e. they were given the same instruction as used in previous tests). Instructions to make global ratings did not appear to explain compensation. However, these instructions caused a slight decrease in perceived channel differences when rated directly. Listeners may have been making
space on the rating scale for differences across the test that were not eventually observed due to compensation. The lack of ability to observe differences between indirectly compared stimuli, even when instructed to do, suggests no role for instructions in the effect seen but also no role for attention, as listeners could not perceive differences even when instructed to pay attention to them. However, attention could be further investigated as a contributor to compensation in future tests outside the scope of this thesis.

Answering Question 3.3 involved determining mechanisms of compensation. The results of previous experiments suggested that two factors were particularly important in causing real-world compensation: the length of continuous listening to the channel and the time between comparisons. In Experiment 2 a preliminary test of the role of longer listening could be carried out (the experiment did not aim to examine this but it was possible to examine this to a limited extent). It was possible to examine ratings that the listeners had made while they had spent approximately 25 s listening within the room/to the loudspeaker to where they had spent approximately 4 minutes listening to the channel continuously. Increased compensation did not occur between approximately 25 s to 4 minutes of listening suggesting that longer listening does not cause compensation. In fact, channel sensitivity appeared to become larger over this period for the room factor suggesting a possible learning of the channel. This is not necessarily incompatible with channel compensation as it is likely that compensation occurs with longer listening but before 25 s, as is common for adaptation or fatigue processes. With extended listening the listener might be expected to become sensitive to the channel once more. As compensation under 25 s was not measured the results do not rule out longer continuous listening as a cause of real-world compensation.

Experiment 3 aimed to investigate Olive et al.’s preferred explanation for real-world compensation and determine the extent to which the time gap between stimuli, when indirectly compared, caused the compensation seen. This would suggest that a type of mechanism that is sensitive to time gaps between comparisons causes compensation. Compensation was moderately and significantly larger with 10 minute gaps for the room factor, compared to no time gaps, showing that the time gaps contributed to compensation but the time gap did not appear to explain compensation effects for the loudspeaker factor. The lack of effect for this factor may be due to ceiling effects (the
room factor used a large amount of the rating scale so there may not have been sufficient
space on the scale for the listener to also show an increased perception of loudspeaker
differences with indirect comparisons). Therefore, the role of time gap in compensation
is confirmed for the room factor but results are inconclusive for loudspeakers. This
result may show a task issue whereby the limited rating scale used in the experiments
prevents proper measurement of compensation and its mechanisms or it may show a
perceptual effect whereby when a factor (e.g. the loudspeaker) dominates perception for
some attribute (e.g. timbral preference) variation in other factors (e.g. the room) on the
same dimension must be suppressed because there is no further room to register this on
a limited perceptual rating scale. This may contribute to any real-world compensation
seen. Overall, it was concluded that enough evidence exists to conclude that a time-
gap sensitive component contributes to real-world compensation. Other non-time-gap
sensitive mechanisms must explain the compensation that remains after the time gap
is removed.

Chapter 3 provided answers to Question 1a. Compensation for loudspeakers and
rooms was shown to occur in real-world listening. Compensation has the effect of
reducing the perception of the channel and creating timbral constancy. According
to preference ratings scales compensation was large to moderate. Research has now
confirmed that compensation occurs with music and speech. The work in this chapter
also contributed to answering Question 1b. Compensation is not due to aspects of
the task that might have biased listeners to intentionally ignore differences in timbre
occurring across the test. Compensation is, at least in part, due to the time gap between
sounds when presented indirectly (a time-gap sensitive mechanism). Compensation
remains after this time gap is removed so it is also due to a non-time gap sensitive
mechanism such as longer listening. A preliminary investigation suggested that it was
also shown that compensation was not due to longer listening between approximately
25 s to 4 minutes but may due to longer continuous listening prior to 25 s. It was
concluded that further work is necessary to establish the mechanisms behind the
time-gap sensitive and non-time gap sensitive components of compensation as it is
not yet possible to determine specific mechanisms behind this. A few mechanisms have
been hypothesised but the role of any existing specific compensation mechanisms in
real-world compensation should be determined.
7.1 SUMMARY OF WORK SO FAR

7.1.3 Chapter 4

Chapter 4 set out a research plan to further determine the mechanisms of real-world compensation. In Experiment 3 causes of compensation (time gaps and instructions) were investigated directly using the real-world paradigm. It was noted that it is possible to determine what type of mechanisms might be behind compensation (i.e. time-gap sensitive or non-time-gap sensitive) but not what specific mechanisms of compensation are at work as too little is known about these. More research is needed to determine whether specific established compensation mechanisms explain the compensation seen in real-world tests. The answers to the chapter questions summarise the work this chapter.

Question 4.1 asked what research is necessary to further determine potential mechanisms of compensation for loudspeakers and rooms when listening to speech and non-speech? A research process was set out in this chapter. This consists of two steps. Step 1 consists of a literature review of existing specific compensation mechanisms from previous psychoacoustic laboratory tests. It was noted that specific mechanisms to be described in the literature review have only been tested in laboratory tests and may not account for the compensation that occurs in real-world listening. To determine whether these mechanisms are likely to play a role in real-world compensation, and the time gap sensitive component of this, 2 criteria where set out: time-gap sensitivity should be established and Questions A-C had to be answered in the affirmative for each mechanism. In order to answer Questions A-C in the affirmative, the mechanism should have a nature that means that it can explain the real-world compensation (Question A), it should be a unique mechanism of compensation, not a manifestation of another, and as a result it will have one or more unique features which can be used to eliminate this mechanism in Step 2 (Question B) and, it should have a time course suitable to explain real-world compensation (Question C). Any compensation mechanism revealed in the literature review that could provide positive answers to Questions A-C and were time-gap sensitive were to be considered a potential mechanisms of real-world compensation.

It was noted that previous tests of specific compensation mechanisms have mainly used speech sounds. If the mechanism is to be regarded a potential cause of real-
world channel compensation generally, and the compensation with music seen in Olive et al.’s (1995) original study, questions A-C should be answered and time-gap sensitivity established for each mechanism when listening to non-speech as well as speech. It was noted that for any given mechanism there may be insufficient information obtained via the literature review to answer Questions A-C for both speech and non-speech listening. In this case the research process prescribed additional laboratory tests to provide this information. Once a list of potential mechanisms is compiled, Research Question 1b would be answered and the aim of this thesis fulfilled. Future research should then proceed to Step 2 whereby each potential mechanism is tested for its actual role in real-world compensation. The real-world paradigm is used and each potential mechanism is eliminated using its unique features established via the literature review (see Question B). This determines the actual role of each potential mechanism in real-world compensation directly (as done in Experiment 3) and confirms the causes of real-world compensation. This step would be necessary to ultimately confirm the role of the potential mechanisms in real-world compensation but was not conducted as part of this thesis. Some of the work to fulfil this is discussed in the next section of this chapter.

Question 4.2 asked about the extent to which the work in this thesis contributes to fulfilling this research process set out. In this thesis potential mechanisms of the time-gap sensitive component of compensation were to be searched for via a literature review and laboratory tests (Step 1 of the research process). It was decided that it was necessary to limit the scope of the research conducted in this thesis so it was decided that Step 1 would only be carried out to determine potential, time-gap sensitive, mechanisms of real-world compensation, thereby explaining the reduction in compensation with the removal of time gaps seen in Experiment 3. Investigation of the non-time-gap sensitive component of compensation is reserved for future work. Further, positive answers to Questions A-C and time-gap sensitivity show potential mechanisms of compensation but these may not contribute or may only contribute a small amount in real-world listening. Actual mechanisms of compensation (Step 2) should be determined but this is reserved for future work. The work in this thesis only investigates the extent to which the psychoacoustic laboratory mechanisms play a role in real-world compensation not the extent of the contribution of any mechanisms that
act in the real world but have not been revealed in psychoacoustic laboratory tests.

Question 4.3 asked how actual compensation mechanisms would be determined. Step 2 of the research process confirms which of the potential mechanisms of compensation actually cause real-world compensation. This involves a process of eliminating each of the potential mechanisms in turn using the unique features to exclude that mechanism from acting in a real-world listening scenario, and measuring the effect on real-world compensation. The work described in this thesis already contributes towards completing this step through answering Question B and identifying the unique features of each mechanism, which can be used in these tests of elimination.

The research questions were not directly addressed in this chapter but a research process was set out to provide further answers to Research Questions 1b and 1c. The remainder of this thesis set out to fulfil part of this research process. It aimed to further answer Research Question 1b and determine potential mechanisms behind the time-gap sensitive component of real-world compensation. A literature review and additional experiments would be conducted to determine mechanisms with the potential to contribute to the time-gap sensitive component of compensation for loudspeakers and room with speech and non-speech listening.

7.1.4 Chapter 5

The aim of Section 1 of this chapter was to begin a literature review of known time-gap sensitive compensation mechanisms from the psychoacoustic literature. From this review mechanisms that are time-gap sensitive and for which Questions A-C could be answered for speech and non-speech listening could be classified as potential mechanisms of the time-gap sensitive component of real-world compensation for loudspeakers and rooms. When one or more such mechanisms is determined Research Question 1b can be answered. It was noted that the literature on compensation mechanisms comes from psychoacoustic laboratory studies from a variety of fields investigating different aspects of speech and non-speech perception therefore Chapter 5 was divided into 4 sections.

Section 1 of the review described compensation mechanisms for spectral variation to
speech caused by vocal tract (VT) differences between talkers and the phonetic context. The review revealed that many perceptual mechanisms may be involved in maintaining constancy in spite of VT variation between talkers. These processes may also cause compensation for the effects of other transmission channels, such as loudspeakers and rooms, which colour the sound in a similar manner to the VT. Further, the phonetic context causes similar spectral distortions as VT variation and compensation for the phonetic context appears to occur by a similar process. Mechanisms that compensate for the phonetic context may also play a role channel compensation. The answers to the chapter questions summarise these mechanisms.

**Question 5.1** asked: a) To what extent does compensation for the spectral effects of the vocal tract occur in speech perception? b) What are the mechanisms behind this? c) What is the time course of this compensation? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

**Question a:** Differences in VT size and shape between talkers mean that the spectral locations of formant frequencies (primary speech cues) are shifted between talkers. This variation is sufficient to cause overlap in vowel space. Accurate perception despite this indicates mechanisms that achieve constancy and possibly VT compensation mechanisms. The extent to which constancy occurs is difficult to assess but it seems adequate to reduce overlap as accuracy of speech identification is good. Some variation between talkers that does not cause overlap is expected to be perceived.

**Question b:** there are a number of mechanisms that may be behind invariant perception across VTs. Phoneme intrinsic mechanisms may be involved. There may be one (target perception) or a many (categorisation) spectral targets, such as formant locations, that remain stable despite VT variation and can be used for identification. This results in the correct category of speech sound being perceived. Distortion may be entirely reduced alongside this categorisation (or it may be perceived to some extent). However, this method may not aid perception where distortion to key targets/sufficient targets occurs and this pushes them to a different region of categorisation. This method of reducing perceived variation may not address the problem of overlap. Categorical Perception (CP) may also be involved in minimising variation in speech cues caused
7.1 SUMMARY OF WORK SO FAR

by VT colouration. CP shifts speech targets that have been varied back to their canonical location (as long as variation is not large enough to signal a completely different category of speech sound, in which case the cue is shifted to the canonical location of that sound). This method of reducing perceived variation also does not appear to address the problem of overlap. Like target perception and categorisation, CP is useful for reducing more minor spectral colouration.

Intrinsic normalisation may occur whereby listeners normalise distorted targets with regard to one or a few other reference cues that are stable within the same phoneme and indicative of the nature of the distortion. Perception of stable target ratios is similar to intrinsic normalisation but appears to involve the perception of targets within a phoneme relative to each other, rather than relative to one or two particular reference cues. These processes can remove vowel overlap. However, the limited information within the phoneme may mean that these processes may not adequately describe the VT frequency response and adequately remove VT colouration. Extrinsic compensation mechanisms provide further information for compensation for VT characteristics. This is a similar process to intrinsic normalisation and target ratio perception but perception is suggested to be relative to the spectral range of the talker as established via a longer segment (a sentence worth) of prior speech: the listener perceives spectral cues within a speech sound relative to spectral cues heard in the prior speech. Like intrinsic normalisation and ratio perception this can remove all distortion and remove vowel space overlap but it appears to give the listener a more complete picture of the VT characteristics of the talker (i.e. the nature of the distortion), because more of the frequency response of the VT can be sampled through hearing an extended sample of prior speech.

Question c: perception of the correct vowel using intrinsic mechanisms (target cues, target ratios and intrinsic normalisation) is immediate or almost immediate. CP and categorisation can be considered phoneme intrinsic mechanisms, occurring immediately upon perception of the sound but there may be a short temporal element to CP. Little data was presented stating the time course of the extrinsic VT compensation processes. Prior experience with the VT is necessary for extrinsic compensation. It may require a sentence-length segment of prior speech. This compensation does not occur when there is more than a 10 second gap between the prior sentence and the test sound.
Question d: little information was provided regarding the extent of these mechanisms within a non-speech listening context. Some of the mechanisms may be specific to compensation for the VT and specific to listening to speech, so they may not occur when listening to other sounds. However, CP and categorisation have been shown to occur with non-speech sounds so these mechanisms may be general auditory compensation mechanisms.

Question e: the extrinsic mechanisms of VT compensation may be considered time-gap sensitive as it is shown that perception relative to the spectral range of prior speech does not occur when the precursor is separated from the test sound. Intrinsic mechanisms are not time-gap sensitive.

Question 5.2 asked: a) To what extent is the phonetic context compensated for? b) What are the mechanisms behind this compensation? c) What is the time course? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

Question a: The prior speech sound causes undershoot due to co-articulation which causes the spectrum of a phoneme to assimilate towards those in its immediate phonetic context. Variation to a phoneme’s spectrum is caused by adjacent speech sounds. This should result in vowel space overlap like that caused by VT variation but like VT variation, this variation does not appear to impede perception. An extrinsic compensation process termed overshoot appears to compensate for the spectral effects of the immediately adjacent phoneme by shifting the perception of each phoneme in a spectrally contrastive direction to the adjacent phoneme. This removes the effects of assimilation and therefore distortion caused by phonetic context. Through this process the overshoot effect appears to enhance spectral differences between adjacent phonemes and enhance spectral change in the speech signal. Thereby fully compensating for the prior phonetic context.

Question b: the physical effect of undershoot is similar to that of speaker variation. In both cases targets are not reached because of limitations of the vocal tract system and there is movement of the cues in a way that is unique to the talker/prior context. In both cases mechanisms are required to shift perception back to canonical targets. In
both cases the spectrum of a prior sound (the immediately prior phoneme for overshoot, the whole prior sentence for VT compensation) is sampled and the perception of the target speech cues is in relation to this, shifting the target away from the prior sound in a contrastive direction. For both effects a relational cognitive process has been suggested. However, Lindblom and Studdert-Kennedy (1967) also put forward a peripheral neural adaptation explanation for overshoot. Target perception (where any stable targets remain), target ratios, intrinsic normalisation, CP and categorisation may aid perception in light of distortion caused by the phonetic context as well as the VT.

**Question c:** little is known about the time course of overshoot from the research described in this section. It appears that only short phoneme-length sounds are needed to bring about this effect. Approximately 80 ms sounds resulted in overshoot in Lindblom and Kennedy’s study. The recovery of effects when the vowel is separated from its consonantal context, are not described. Its whole time course is likely to be short and it is probably also time-gap sensitive.

**Question d:** like compensation for the VT, overshoot may be a specific process that has evolved to remove colouration caused by the phonetic context when listening to speech. It is not clear whether this process can occur to compensate for the effects of other channels when listening to non-speech sounds. Likewise intrinsic methods such as target perception, that might occur to provide stability in light of the problems of co-articulation may also be speech specific but CP and categorisation are not.

**Question e:** The overshoot effect is likely to expire with a silent gap between a target vowel and its phonetic context so it is probably time-gap sensitive but this was not confirmed.

**Question 5.3** asked: a) To what extent is the effect of transmission channels other than the vocal tract (e.g. loudspeakers, rooms) compensated for in speech perception? b) What are the mechanisms behind this? c) What is the time course of effects? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

No direct evidence was presented to answer Question 5.3. The tests described in this
section involved compensation for speech specific sources of spectral colouration. It may be that the mechanisms described evolved in response to speech specific problems so compensation may not occur to remove the effects of other channels or when listening to sounds other than speech. Further, other channels may cause less extreme distortion so the extent of compensation caused by the same mechanisms may be reduced.

**Question 5.4** asked: a) To what extent does compensation for transmission channels occur when listening non-speech sounds? b) What are mechanisms are behind this? c) What is the time course of effects? d) Is compensation for the transmission channel spectrum a general auditory process? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

The extent of compensation when listening to non-speech sounds was not investigated as most of the research described involved an entirely speech context. Some perceptual mechanisms that occur when listening to speech may be speech specific and a speech mode of perception may be necessary to instigate the compensation described here. CP and categorisation have been shown to occur when listening to non-speech sounds.

**Section 2**

Section 1 investigated compensation for spectral colouration caused by the VT and the phonetic context. The studies in Section 2 of the literature review aimed to determine perceptual mechanisms that compensate for colouration caused by transmission channels more generally, as well as provide further information on the VT/phonetic context compensation mechanisms described. The answers to the chapter questions summarise these mechanisms.

**Question 5.1** asked: a) To what extent does compensation for the spectral effects of the vocal tract occur in speech perception? b) What are the mechanisms behind this? c) What is the time course of this compensation? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

Summerfield and colleagues suggested that a perceptual process by which prior listening
to a noise masker causes reduced masked thresholds of signals may be involved in compensation for channel colouration. They noted that this process would create compensation for the channel effect and relatively enhanced perception of the source. A flat spectrum harmonic complex or noise precursor was filtered by an inverse vowel shaped filter. Compensation was demonstrated when the same flat spectrum sound was presented immediately after the precursor as a test sound with the filtering removed and the vowel filtered for in the precursor was heard in the test sound. This effect is termed the enhancement effect and demonstrates that during presentation of the precursor, channel compensation builds up and is applied to the test sound. The enhancement effect may be a mechanism of VT compensation as this effect appears to enhance spectral change in a source relative to the static channel. This effect is monaural and does not occur when the precursor and test sound are presented to different ears. Tests by Watkins and colleagues also showed that flat noise precursors filtered by inverse vowels caused the enhancement effect in test stimuli. Watkins showed again that the effect was monaural and showed that it was not sensitive to perceived direction change between precursor and test sound. Both Summerfield’s work and that of Watkins et al. suggest that a peripheral adaptation explanation for enhancement is likely based on its monaural nature and short time course. Simple adaptation or adaptation of suppression may be involved. The effect is quick to onset 150 ms and recovery with 500 ms. The mechanism may be considered time-gap sensitive as it recovers with approximately 500 ms of silence and therefore will not apply compensation for a precursor to a sound where there is sufficient silent gap between them. The enhancement effect appears not to be speech specific as there is some evidence of this occurring with sine-tones and noise in masking studies. However, studies specifically investigating this effect as a means of channel compensation have used speech stimuli somewhere in the test so the general auditory nature of enhancement is not confirmed. Speech anywhere in the test may elicit a speech mode of listening.

Watkins also tested the enhancement effect with real-speech sentence-length precursors filtered with inverse vowels (rather than static noise or harmonic complex precursors). Shifts in the perception of test vowels occurred in the same way as before but were from an apparently different source. These shifts occurred crossaurally showing that a central process must cause the shifts. This central compensation effect was termed the
7.1 SUMMARY OF WORK SO FAR

spectral compensation effect to distinguish it from the enhancement effect. Watkins and Makin’s (1994) study showed directly that the spectral compensation effect was behind the extrinsic VT compensation process described in Section 1. The spectral compensation effect appears to be more suited to compensation for the VT than enhancement. In VT compensation studies it was apparent that the spectral cues within a whole precursor sentence were a factor in the nature of shift so this further suggests that the spectral compensation effect is in response to the LTAS of a longer sentence rather than the immediately prior phoneme. A sentence-length precursor may be necessary but another study reported the spectral compensation effect occurring with a short sound following the test sound, so a long precursor may not be necessary to bring about the effect (Watkins and Makin 1994). The recovery time appears to be longer than enhancement. It continued with over 160 ms, where the enhancement effect had expired. If the mechanism is the same as that in Ladefoged and Broadbent’s (1957) study it may break after 10 seconds. It is therefore time-gap sensitive and will not apply compensation for a precursor to a sound where there is a silent gap between them. The mechanisms underlying the spectral compensation effect were not uncovered in this section but it was noted that real-speech may instigate the effect or spectral variation in the precursor may be necessary. The backwards-acting effect appears to rule out a peripheral cause, as does the fact that the mechanism is centrally produced. The authors put forward a memory hypothesis due to the effect working backwards and appearing to calculate LTAS over a longer than phoneme-length period. The spectral compensation effect has so far only been tested where speech is present either in the precursor or test sound (it had not occurred previously where noise and harmonic complexes were used). It may require speech and it may be speech specific. However, it may only require spectral variation in the precursor (which was not present in tests showing only enhancement). If it is found that this mechanism simply requires spectral variation in the precursor then this may be a general auditory process. Other features of the spectral compensation effect, such as its direction sensitivity and time course, appear to be appropriate for channel compensation. Watkins concluded that the spectral compensation with its longer time course is more likely to cause channel compensation than the enhancement effect, as was suggested by Summerfield et al.

Question 5.2 asked: a) To what extent is the phonetic context compensated for? b)
What are the mechanisms behind this compensation? c) What is the time course? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

Overshoot was described as a mechanism that primarily compensates for the phonetic context. This section reveals that the enhancement effect is similar in nature and time course to the overshoot effect and may be the cause of overshoot, although this was not specifically suggested by the authors of the enhancement work. Both appear to provide perceptual shifts that show compensation for the spectrum of the immediately prior sound, which has the effect of pushing the currently heard sound away from the spectrum of the prior sound (thus pushing a phoneme affected by undershoot to its canonical target) and enhancing change. Further, peripheral neural adaptation explanations have been offered for both. The enhancement effect appears to be well suited to be a mechanism for compensation of the phonetic context as its time course is short. The role of the spectral compensation in compensation for phonetic context is not clear. The data in this section showed that the time course appears be longer than enhancement and compensation might be for the LTAS of a longer sentence. In which case, in running speech, this effect will remove the effects of the channel but not the prior phonetic context. The time course, time-gap sensitivity and general auditory nature of these effects were discussed in the answer to Question 5.1.

**Question 5.3:** asked a) To what extent is the effect of transmission channels other than the vocal tract (e.g. loudspeakers, rooms) compensated for in speech perception? b) What are the mechanisms behind this? c) What is the time course of effects? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

The work presented does not involve the examination of compensation for the phonetic context or VTs specifically but compensation for filtering more generally when listening to speech. The filtering in this work is more representative of filtering caused by generic channels such as loudspeakers and rooms because it is wideband. However, the filtering was for inverse vowel spectra—Summerfield’s work consisted of notch filtering to represent formant peaks and Watkins’ studies used smoother filter transfer functions but still *inverse-vowel shaped* containing spectral peaks. This filtering may therefore
be considered more extreme than that caused by loudspeakers and rooms. However, it is expected that the enhancement effect and spectral compensation may both play a role in compensation for this milder filtering. The time course of enhancement is quick but that of spectral compensation may be longer. As mentioned above, a longer time course may be more suitable for compensation for static channels as the LTAS will be used in compensation. The direction sensitivity of the spectral compensation effect may mean that it is less suitable for compensation for the room. Mechanisms behind these effects, time course, time-gap sensitivity were discussed in the answer to Questions 5.1 and 5.2.

**Question 5.4** asked: a) To what extent does compensation for transmission channels occur when listening non-speech sounds? b) what are mechanisms are behind this? c) What is the time course of effects? d) Is compensation for the transmission channel spectrum a general auditory process? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

All work described in Section 2 used speech sounds somewhere in the test and this may have brought about a speech mode of listening. This limits conclusions regarding the general auditory nature of effects. An enhancement-like effect is shown to occur in non-speech tests with noise and tones (i.e. masking tests), but these may elicit different effects to the *channel compensation* effects seen here. There was also strong evidence that it occurs via low level neural adaptation, so the balance of the evidence currently suggests that it is not speech specific. The fact that the spectral compensation effect does not occur with noise suggests speech specificity but the spectral compensation effect was shown not to be unique to natural-sounding speech sounds occurring when speech precursors were reversed. It was concluded that it may simply require spectral variation and may be a general auditory process.

**Section 3**

The last section of the review set out to describe studies that have investigated compensation in non-speech perception. However, it was noted that studies that have aimed to investigate compensation in non-speech perception have used speech in the
7.1 SUMMARY OF WORK SO FAR

test in most cases so do not conclusively show that these compensations mechanisms are general auditory process. A speech mode of listening may be engaged. Therefore, these studies primarily provide further information regarding the above mentioned compensation process when speech is present. However, some information regarding the possible general auditory nature of these mechanism was provided. The answers to the chapter questions provide a summary of the findings for this chapter.

**Question 5.1** asked: a) To what extent does compensation for the spectral effects of the vocal tract occur in speech perception? b) What are the mechanisms behind this? c) What is the time course of this compensation? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

Overshoot and the enhancement effect have been suggested to contribute to channel compensation, including VT compensation. Studies that measured the overshoot effect with non-speech sine tone precursors were described in this section. Because speech was present in the test sounds, these studies do not conclusively show overshoot in a non-speech context. However, these tests provide further information about the overshoot effect and specifically, further evidence of the similarity of the enhancement effect and overshoot. The overshoot studies presented in this section show that a monaural mechanism contributes to overshoot and confirmed the short time course of overshoot. The nature and time course matches the enhancement effect, so this effect is likely to be behind overshoot. Because both have the same nature, enhancing spectral change relative to a short phoneme-length precursor, and time course, it was concluded that overshoot and enhancement may be regarded as examples of the same mechanism. It therefore appears that both may be peripheral auditory mechanisms and as well as providing compensation for the phonetic context overshoot/enhancement may provide general channel compensation, including compensation for the VT via enhancing the changing sound source relative to the channel.

Section 3 of the literature review presented good evidence for the spectral compensation effect occurring with non-speech precursors. This work showed a new *acoustic histories* effect, which appears to be a manifestation of the spectral compensation effect occurring with sentence-length non-speech precursors. Shifts in the perception of speech tests
sounds in the opposite timbral direction to the mean spectrum of the precursor were seen. Evidence for this effect being the same as the spectral compensation effect derives from the fact that, like the spectral compensation effect, a study has directly shown the acoustic histories effect to cause the compensation for the VT described in Section 1 of the review. This implies that compensation for the VT, the spectral compensation effect and the acoustic histories are all caused by the same underlying mechanism. As for the spectral compensation effect, the acoustic histories effect is shown to be central and it was confirmed in this study that compensation is relative to the LTAS of a whole sentence length precursor, rather than the immediately adjacent spectrum. Therefore this effect is now considered an example of the spectral compensation effect. Acoustic histories studies therefore shed light on the spectral compensation effect confirming that this produces compensation for LTAS: shifts in the perception of test sounds in the contrastive direction to the mean spectrum of a longer than phoneme-length precursor occur. This compensation for the LTAS strongly suggests a mechanism that compensates for the channel, including the VT, as it samples a static spectrum over a period long enough to gain a picture of the channel response and applies this in compensation. Further, the spectral compensation effect has now been shown to directly contribute to compensation for the VT by two studies.

Further information on the time course of enhancement/overshoot was gained via work within this section. It was shown to have a short phoneme-length time course. Given that the acoustic histories studies show the spectral compensation effect, these studies provide further evidence on the time course of the spectral compensation effect. The time course is confirmed to be longer than that of enhancement. It appears to onset with as little as a phoneme-length precursor (as seen in overshoot studies), but where precursors are longer than the phoneme, compensation is for LTAS of a sentence-length precursor. It is not clear what the window length of the onset is, as maximum onset was not observed. The offset of the effect is slower than the enhancement. The effect lasts over at least 1.3 s. The offset window appears to be 10 s as shown in Ladefoged and Broadbent’s (1957) work. The slow offset of this mechanism means that it can effect following sounds but this does not occur with silence between the precursor and test sound so can be considered time-gap sensitive.

The underlying mechanism behind the enhancement effect appears to be peripheral
adaptation as the time course is short and effect monaural. The mechanisms behind
the spectral compensation effect are suggested to be timbral memory and/or stimulus
specific adaptation occurring at central sites. It was also hypothesised that MOC noise
reduction process may contribute either to enhancement or spectral compensation.
Little new evidence was provided to show that enhancement/overshoot effect or
the spectral compensation effect occur in an entirely non-speech context as speech
perceptions were elicited in most tests. Studies only using non-speech are inconclusive
of these effects. Further discussion of speech specificity is given in the answer to
Question 5.4.

Given that overshoot and enhancement are the same, tests that show overshoot with
non-speech can shed light on the workings of enhancement/overshoot with non-speech.
Studies by Holt et al. and Lotto et al. tested overshoot using the traditional paradigm
with non-speech precursors. Overshoot occurred with these sine tone precursors but
the similarity of these precursors to speech and speech in the test sound meant that
these studies do not confirm that overshoot/enhancement is a non-speech process.
The evidence so far that enhancement/overshoot is a general auditory process was
assessed. It appears to a) be caused by an obligatory peripheral auditory mechanism
that will occur with any sound; b) masking studies show an enhancement-like effect
occurs in an entirely non-speech context with noise and static sine tones and c) non-
speech precursors bring about the effect (though speech is present in test sounds).
It is therefore concluded that enhancement/overshoot is probably not speech specific
but this must be tested further in a study that uses only non-speech sounds because:
a) no direct physiological evidence has been found to show that enhancement is an
obligatory peripheral auditory process; b) masking studies show an enhancement-like
effect but are significantly different to the enhancement studies by Summerfield which
are claimed to show channel compensation (e.g. they may show two-tone suppression
rather than enhancement), c) the presence of speech in the test sounds may be sufficient
to instigate a speech mode of listening. If enhancement/overshoot is a mechanism to
reduce the impact of the phonetic context it would not be surprising to find that this
effect is speech specific and has evolved to solve a speech specific problem.

Given that the acoustic histories effect is a manifestation of the spectral compensation
effect, acoustic histories studies appear to provide good evidence of the spectral
compensation effect occurring with non-speech precursors. However, this is not conclusive evidence of the effect being a general auditory mechanism because speech test sounds were used and these may be sufficient to bring about the effect. Three tests using a test method similar to that used by Watkins but only non-speech sounds were described in section 3. These studies showed some evidence of channel compensation by an extrinsic mechanism. The studies did not adequately test whether any shifts were centrally produced, so it is not clear whether this showed an enhancement or spectral compensation effect causing this compensation, they used unnatural stimuli meaning that any absence or reduction of shift could be due to this, or they contained methodological differences and differences in the aims, which prevented a conclusion that spectral compensation effect occurred with non-speech.

Question 5.2 asked: a) To what extent is the phonetic context compensated for? b) What are the mechanisms behind this compensation? c) What is the time course? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

The overshoot effect was investigated further in the studies described in this section. Enhancement was shown to be likely to cause overshoot as overshoot was confirmed to be monaural and has a short time course. A central effect may also be involved in causing overshoot but this is likely to be a manifestation of the spectral compensation with short precursors. Enhancement/overshoot appear to contribute to compensation for the phonetic context by compensating for the spectrum of the immediately prior phonemes worth of sound. The exact mechanisms behind this process (e.g. peripheral adaptation) were discussed in the answer to Question 5.1. The spectral compensation is unlikely to contribute to compensation for the phonetic context as its time course is too long. In running speech compensation for the LTAS of a longer segment will be applied rather than the immediately adjacent phoneme. The time course and time gap sensitivity of the enhancement/overshoot and the spectral compensation effect were discussed in the answer to Question 5.1. A discussion of the extent to which these effects occur in an entirely non-speech context is saved for Question 5.4.

Question 5.3 asked: a) To what extent is the effect of transmission channels other than the vocal tract (e.g. loudspeakers, rooms) compensated for in speech perception?
b) What are the mechanisms behind this? c) What is the time course of effects? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

As discussed before in Sections 1 and 2 of the review, studies on overshoot/enhancement may be uncovering a speech specific mechanism. However, the case is strong that enhancement/overshoot is a general auditory process. Enhancement has been frequently discussed as a mechanism of channel compensation, and the evidence so far implies that this (and therefore the overshoot effect) can contribute to compensation via enhancing change relative to any channel. Therefore, it is considered to be a mechanism of channel compensation. However, work should confirm that the nature of the effect is appropriate to remove the channel and that it occurs with non-speech sounds. The spectral compensation mechanism appears to be a general auditory process that compensates for LTAS, so may remove effects of any channel. However, it was observed that further entirely non-speech tests should be conducted to confirm this mechanism can occur with non-speech sounds. Direction sensitivity may still be an issue with this effect when compensating for room reflections. However, until this is confirmed this mechanism is regarded has having the the potential to provide compensation for colouration caused by the room, as well as the loudspeaker. The time course and time-gap sensitivity of these effects were discussed in the answer to Question 5.1. A discussion of the extent to which these effects occur in an entirely non-speech context is saved for Question 5.4.

**Question 5.4** asked: a) To what extent does compensation for transmission channels occur when listening non-speech sounds? b) what are mechanisms are behind this? c) What is the time course of effects? d) Is compensation for the transmission channel spectrum a general auditory process? d) Can the mechanisms behind this compensation be considered general auditory processes? e) Can they be considered time-gap sensitive?

Despite the main aim of this section being to determine compensation mechanisms in tests which do not use speech, little information was obtained regarding this. This is because almost all tests purporting to test the speech specificity of these mechanisms used speech in the test sounds. This may instigate a speech mode of perception. The
work in this section showed that overshoot occurs with non-speech precursors. This is suggestive of a non-speech mechanism behind overshoot/enhancement, but the use of sine tone transition precursors and speech test sounds mean this is not confirmed via this work. However, the evidence presented across all 3 literature review sections suggests that overshoot/enhancement is likely to be a general auditory process. It was concluded that an entirely non-speech study is needed to confirm this.

Firm conclusions regarding the spectral compensation effect with non-speech listening could also not be made. Holt’s work was the first to test this with a non-speech precursor and shows that the effect occurs. Spectral variation rather than speech in the precursor may be all that is necessary to bring about the effect. A few tests using only non-speech sounds showed that at least either enhancement or spectral compensation occur with non-speech but which mechanism could not be confirmed because the central nature of the effect was not confirmed, unnatural stimuli were used or there were too many differences in aims and methodology to confirm the presence of spectral compensation effect. It was concluded that further work examining these mechanism with non-speech sounds should be carried out as part of this thesis.

Section 4

In the literature review the nature and time course of each of the mechanisms was described in as much detail as available. In section 4 this information was used to determine which of the mechanisms are time-gap sensitive and whether Questions A-C could be answered for each mechanism, so that potential mechanisms of the time-gap sensitive component of real-world compensation could be determined. No chapter questions were addressed in this section but the answers to Questions A-C and time gap sensitivity for each of the mechanisms described in the literature review is summarised.

In answering question A the nature of the mechanisms was determined and whether or not the mechanisms had the nature to cause the contraction of ratings and the shifts towards the perceptual midpoint seen in real-world listening tests. All of the mechanisms (target perception, categorisation, CP, intrinsic normalisation, target ratio perception, extrinsic compensation for the VT, overshoot, enhancement, the spectral compensation effect, peripheral neural adaptation, central neural adaptation,
the acoustic histories effect, timbral memory) except timbral memory were regarded as *compensatory* and having the nature capable of causing the contraction of ratings seen in real-world tests. Timbral memory was concluded not to be likely to cause the required shifts in perception, but to result in an increase or decrease in perceptual noise. Therefore, this was ruled out as a potential mechanism of real-world compensation due to its apparent non-compensatory nature.

In answering Question B it was determined whether or not the mechanisms had distinguishing features that showed that they were a unique mechanisms and not a manifestation of another (and also so that they can be distinguished for the purposes of elimination in Step 2 tests). The evidence from the literature pointed to a number of mechanisms that had been investigated separately as being the same. It was shown that of the mechanisms with the nature to cause compensation, the unique intrinsic mechanisms are: CP, target perception, categorisation, intrinsic normalisation. The unique extrinsic mechanisms are the spectral compensation effect and the enhancement effect.

Question C involved confirmation that the time course of these unique mechanisms is appropriate to explain real-world compensation. It was confirmed that the spectral compensation effect and the enhancement effect have the time course to explain this because the are extrinsic. This means that they require time spent listening to onset. They do not onset immediately and therefore they are more likely to have onset (or to have become large) with the indirect comparison condition in real-world tests, which involves longer listening periods, than the direct comparison condition. Therefore, there is likely to be more channel reduction with indirect comparisons. This may explain increased channel compensation in real-world listening. However, it was noted that this is only true for the spectral compensation effect the enhancement effect with its quick onset is likely to reach maximum with both direct and indirect comparisons. Intrinsic mechanisms would definitely onset equally with both direct and indirect comparisons, because they occur immediately, so do not have the time course to explain real-world compensation. Therefore intrinsic mechanism are ruled out as explanations for real-world compensation because they act equally with direct and indirect comparison conditions.
7.1 SUMMARY OF WORK SO FAR

While enhancement may not cause compensation that results from longer listening, both extrinsic mechanisms may cause compensation by enhancing differences between stimuli heard close in time but not when they are separated. Once the compensation has onset a change in the channel would be enhanced due to this compensation mechanism. Both the enhancement effect and the spectral compensation effect may reach complete onset with direct and indirect conditions. This would result in any immediate change of channel being enhanced. However, both would offset with the time gaps involved with indirect comparisons, but not direct comparisons. This lack of enhancement of adjacent differences with indirect comparisons can explain the reduced perception of timbral differences with indirect comparisons. The extrinsic mechanisms are classed as time-gap sensitive because the can bring about reduced timbral differences (compensation) via a lack of enhancement where there are time gaps.

It was concluded that of all the mechanisms reviewed only the enhancement effect and the spectral compensation effect have the potential to explain the time-gap sensitive component of real-world compensation for loudspeaker and rooms. However, it was noted that these effects have not been confirmed to cause compensation in non-speech listening so further experiments were required to show this.

7.1.5 Chapter 6

In Chapter 5 it was concluded that the enhancement effect and the spectral compensation effect are potential mechanisms of the time-gap sensitive component of real-world compensation when listening to speech. However, it was not clear if these occur when listening to non-speech so these mechanisms may not be general mechanisms of real-world compensation for loudspeakers and rooms. This chapter aimed to determine whether these mechanisms arise as part of a speech mode of perception or whether they also occur with non-speech sounds. These experiments were necessary to confirm potential mechanisms of the time-gap sensitive component of compensation for loudspeakers and rooms when listening to non-speech, including the music used in Olive et al.’s original test. The answers to the chapter questions summarise the work in this chapter:
**Question 6.1** asked: ‘to what extent does the work in this chapter show that *any* of the specific compensation mechanisms cause compensation in non-speech perception?’

In Experiment 4 evidence that at least one of the extrinsic mechanisms (enhancement and spectral compensation) causes compensation with non-speech was seen as shifts indicative of these mechanisms were seen in an entirely non-speech tests with both music and noise, including the same music used in Olive et al’s original experiment. It was therefore shown that either enhancement and/or spectral compensation occurred with both music and noise. The shift was a similar size (approximately 22% shift) for all programme items but the exact mechanisms behind the shift for each programme item may be different.

**Question 6.2** asked: ‘exactly which of the compensation mechanisms result in this compensation with non-speech?’ Experiment 5 tested whether the enhancement effect, the spectral compensation effect or both occurred with the music and noise programme items. Experiment 5 showed that, as expected, noise appears to only elicit a monaural compensation effect. This is strongly suggestive of the enhancement effect occurring with noise and no contribution from a central mechanism. It was noted that it remains possible that the effect was caused by a central direction sensitive process but that this is unlikely as previous research by Watkins et al. found only a monaural effect and no direction sensitivity with noise precursor stimuli. With music a central component to the shifts was seen. This shows the spectral compensation mechanism working in a non-speech context. A monaural effect, indicative of the enhancement effect, also appeared to contribute to compensation with music. Again, further testing of direction sensitivity could be undertaken to confirm the existence of this monaural effect but the balance of evidence suggests that the enhancement plays a role in compensation with music.

**Question 6.3** asked: ‘are any of these mechanisms potential mechanisms of compensation for loudspeakers and rooms in real-world listening to non-speech?’

The results from Experiment 5 confirm the existence of the enhancement effect and the spectral compensation effect when listening to non-speech. Upon finding these effects it could be assumed that the criteria for concluding that these effects are potential mechanisms of the time-gap sensitive component of compensation for loudspeakers...
and rooms in real-world listening were met. This is because these mechanisms have already been shown to be unique, and have a nature and time course that is appropriate to explain real-world compensation and the time-gap sensitive component of this with speech listening. There is nothing about non-speech listening that is expected to change the features or time course of these mechanisms to the extent that this is not longer the case. (i.e. the criteria of time gap sensitivity and answering Questions A-C in the affirmative were automatically met upon finding these mechanisms with non-speech).

It is therefore concluded that the enhancement effect and the spectral compensation effect are potential mechanisms of the time-gap sensitive component of compensation for spectral colouration caused by loudspeakers and rooms in real-world listening with speech and non-speech. But it appears that the central spectral compensation does not have a role in compensation with noise in laboratory tests (i.e. it is not a potential mechanism of compensation for loudspeakers and rooms when listening to noise). Therefore, it will not have a role in real-world compensation with noise. Both the spectral compensation effect and enhancement effect have the potential to play a role in real-world compensation with music.

It was concluded that Experiments 4 and 5 aimed to confirm that extrinsic mechanisms—the enhancement effect and the spectral compensation effect— which have the potential to explain the time gap sensitive component of real-world compensation do not arise as part of a speech mode of perception but occur with non-speech as well as speech. Both effects were shown to occur in laboratory tests with the same music used in Olive et al’s original experiment and noise but the spectral compensation effect did not occur with noise. Research Question 1b has now been answered for both speech and non-speech listening. Research Question 1b was answered but it was acknowledged that real-world tests could be conducted to confirm the role of these mechanisms in such compensation (i.e. Step 2a of the research process can be conducted) and further tests could investigate potential and actual non-time-gap sensitive mechanisms, which are likely to have a large role in real-world compensation with both music and noise (Steps 1b and 2b can be conducted). The work in this chapter has provided stimuli that could be used in future tests of compensation under a variety of different conditions.
7.2 Future Work

There has been relatively little research into the compensation for spectral colouration caused by general transmission channels. There has been even less into compensation for the spectral effects of loudspeakers and listening rooms and the mechanisms that may occur to cause this during real-world listening. Further, and more surprisingly, there has been relatively little research into timbral-constancy as a general phenomenon in hearing yet a great deal of research into colour-constancy in vision. These research fields would greatly benefit from further work.

This thesis showed a number of specific mechanisms that appear to cause compensation for spectral colouration caused by transmission channels. Some of these mechanisms appear to explain compensation that occurs in the real world. These appear to explain why larger timbral differences are perceived between loudspeakers and listening rooms when listening in a laboratory scenario compared to listening in the real world.

The work in this thesis contributed to explaining part of the reason for real-world compensation: that time-gap sensitive extrinsic compensation mechanisms might cause enhancement when stimuli are presented side-by-side in the laboratory but not in the real-world. Further work is necessary to explain other perceptual contributions to real-world compensation. Further, the work in this thesis has developed a list of potential mechanisms that might contribute to real-world compensation, but it has not been confirmed that these mechanisms actually do contribute to real-world compensation. Further work must be done to examine the contribution of the potential mechanisms to real-world compensation.

The future work that is discussed here has been divided into 3 main areas:

Section 7.2.1, Further work to establish specific channel compensations mechanisms: There is still considerable research to be carried out investigating specific compensation mechanisms, such as the spectral compensation effect and the enhancement effect and their role in channel compensation generally.

Section 7.2.2. Further work to establish the actual role of potential mechanisms in explaining the time-gap sensitive component of real-world
compensation: This thesis determined potential mechanisms but further research is necessary to determine the actual role of these mechanisms in the time-gap sensitive component of real-world compensation. Future work to determine whether the enhancement effect and the spectral compensation effect explain real-world compensation (in particular the time-gap sensitive component of this) will be described.

Section 7.2.3. Further work is necessary to determine whether mechanisms other than the enhancement effect and the spectral compensation could explain the time-gap sensitive component of compensation seen in this work: Memory is discussed as a possible explanation.

7.2.1 Tests to further determine mechanisms of channel compensation

The literature review involved investigating specific channel compensation mechanisms that may play a role in real-world compensation. However, the extent to which these cause channel compensation was not always clear. Throughout the discussion of the specific compensation mechanisms it was noted that a number of questions remain unanswered and further work should be conducted. In most cases additional experiments are necessary to do this but information may also be gained by further searches of the literature.

Future work to determine the nature and attributes of these compensation mechanisms is discussed here. This discussion is divided into discussions of work concerning intrinsic and extrinsic mechanisms. Compensation by intrinsic mechanisms is immediate and acts on sounds when heard in the laboratory (i.e. when making direct comparisons in real-world compensation studies). They also act when listening in the real world (i.e. when making indirect comparisons in real-world studies). There is no evidence to suggest that they elicit more compensation when listening in the real world. Therefore, the intrinsic mechanisms were shown not to contribute to real-world compensation. However, it may be of interest to research these mechanisms further, as they still contribute to channel compensation generally. It is apparent that any compensation for colouration caused via intrinsic mechanisms is not complete, as the channel is
heard when listening over short listening periods (i.e. in the laboratory). This leaves room for further compensation to occur via extrinsic mechanisms (i.e. in the real world). Those investigating real-world compensation should prioritise work clarifying the nature and attributes of extrinsic mechanisms, rather than intrinsic mechanisms. However, intrinsic mechanisms may reduce channel effect considerably compared to if they did not occur. Future work to clarify the role of each of the mechanisms discussed in the literature in channel compensation is described in brief.

**Target Perception**

**Nature:** It was concluded that target perception (e.g. the perception of specific formants in speech perception) aids auditory object recognition but it is not entirely clear whether this causes channel compensation. The target perception hypothesis suggests that one or two targets that are rarely affected by distortion allow the listener to correctly identify an auditory object despite distortion caused to other aspects of the spectrum. This appears to be a process of redundancy. Identification occurs in spite of distortion. It appears that perception of the distortion may still occur and it may not be greatly reduced. Further work should be done to determine the extent to which target perception not only aids object identification but also reduces perceived colouration to that object.

The correct identification of auditory objects (e.g. vowels) is usually used to measure stability that results from target perception. This does not measure the perception of the channel and it is possible that there is some compensation of channel. It may be that the listener is less sensitive to channel effects when targets are available compared to situations where some or all of the key targets are distorted. The correct identification of auditory objects cannot be used to test this hypothesis, as the target perception hypothesis proposes that correct identification occurs regardless of the level of perceived distortion. The extent to which the channel is perceived, rather than the extent to which the object is correctly identified, must be measured to shed light on whether there is a reduction in the channel effect alongside accuracy of identification. Tests could be run whereby the availability of targets is manipulated and the perceived contribution of the channel to perception is measured. Sounds with these varying numbers of targets removed may be played through different channels, this creates timbral differences in
these sounds. The listeners could be asked to rate timbral difference thereby measuring the effect of the channel. This can be done with speech (e.g. by removing formants or other targets) and non-speech (by manipulating other features).

**Time course:** Target perception has been classified as an instant mechanism as no prior listening beyond the test sound is required for this effect to occur. However, this process may take milliseconds to occur. It may be interesting to further probe the exact time course of *instant* intrinsic mechanisms using reaction times or physiological measurements. Investigating time course can shed light on the specific neural mechanism behind an effect. However, in terms of establishing the nature of the effect, it may not be particularly interesting to investigate time course, as it does not dictate when compensation will and will not occur in the same way as time course does for extrinsic compensation mechanisms.

**Magnitude:** If there is any reduction in channel effect due to target perception then the magnitude of this needs to be established. As with any instant compensation, the fact that the channel effects can still be heard over short listening time courses shows that compensation by this mechanism is not complete. However, the channel effect may be reduced significantly compared to a situation where this mechanism does not occur. It is difficult to measure the magnitude of instant sources of compensation as they are always present. There is no uncompensated baseline against which to measure compensation. It may be possible to develop methods of measuring perception before and after any compensation that occurs almost instantly. Reaction time measures or direct physiological measures may be used to measure quick acting effects.

To measure the size of instant compensation, perception can be compared to that expected based on objective measures. If perception is reduced compared to that suggested by an objective measure compensation may be said to have occurred. However, in this case it is difficult to distinguish compensation mechanisms from limitations of the hearing system. A limitation of the hearing system might be a process whereby (through evolution) the hearing system has adapted to become less sensitive to certain aspects of sounds in the environment to improve accuracy of perception of the most relevant auditory objects. Compensation on the other hand is a process that allows humans to adapt quickly to a variety of situations (e.g. different rooms). It
may be difficult to distinguish whether a hearing system removing a channel or room is exhibiting an adaptive (evolutionary) trait or a situational compensation process. For example, target perception appears to be a mechanism by which the hearing system is only sensitive to particular cues and this is the case in all situations. The cues to which it is sensitive do not vary in different situations, so this might not be regarded as a compensation mechanism, but more specifically an adaptive mechanism. Further tests should aim to determine the difference between limitations of the hearing system that aid perception in all listening and compensation mechanisms that apply different limitations in different situations, as this research would be useful to understanding the process of channel compensation and timbral constancy generally. More specifically, future research should examine the extent to which target perception is a compensation process or represents limitations of the hearing system.

**Mechanisms:** The research in this review did not shed light on the mechanisms behind target perception but it was noted that combination sensitive neurons may exist to detect targets (Sussman 1989, Sussman et al. 1997) which are similar to those known to occur in vision for the detection of lines and edges (Hubel and Wiesel 1959, Shapley and Tolhurst 1973). If target perception can be considered a compensation mechanism further work should be done to determine the mechanisms behind this.

**Compensation in non-speech perception:** Targets may play a similar role in identifying non-speech sounds in spite of distortion. Instead of formant frequencies other stable spectral features of a sound may be used in identification.

**Categorisation**

**Nature:** This mechanism appears to be the same process as is involved in target perception but a number of undistorted targets lead to the correct perception of the sound, rather than just one or two as is proposed with target perception. The availability of many targets means that these can be traded where distortion affects some targets as long as sufficient targets remain to categorise the sound correctly. As with target perception the channel distortion may still be perceived to a large extent and further research may be necessary to confirm that this process does reduce the channel as well as aiding auditory object recognition. As with target perception,
7.2 FUTURE WORK

studies that aim to measure the perception of the channel rather than identification performance may be used to investigate this further.

**Magnitude:** If categorisation is a compensation mechanism then the magnitude of this needs to be established. As with target perception, and any other intrinsic immediate mechanism, it is difficult to measure the magnitude of compensation as it is difficult to measure an uncompensated baseline perception. Perception can be compared to that expected based on objective measures or methods may be developed to measure the very rapid adjustment that occurs with instant mechanisms.

**Time course:** As for target perception, physiological measures or reaction time may be used to establish the time course of this quick acting intrinsic mechanism. The time course may be of interest for establishing underlying brain mechanisms but the time course information is unlikely to be of interest in establishing the scope of the effect.

**Mechanisms:** The process of categorisation is a general auditory and general cognitive process. There is a large amount of literature on categorisation mechanisms in cognitive psychology, this can be investigated further to determine the potential of this mechanism as a method of channel compensation and the mechanisms behind this (see Holt and Lotto (2010)). The literature on trading relations in speech (Repp 1981), whereby target cues can be traded if distortion affects some but not others is also likely to be informative of categorisation as a means of reducing distortion, as this process appears to be similar to categorisation.

**Non-speech perception:** Categorisation is a general cognitive process and it is likely to occur in the identification of most objects including non-speech auditory objects. Categorisation has been the subject of much cognitive psychology research but more studies into categorisation of speech and non-speech are likely to be useful (see Holt and Lotto (2010)).

**Target ratios**

In Section 5.4.1 it was concluded that the process of timbral constancy via perceiving target ratios may be the same process as intrinsic compensation. Therefore the role of target ratios in compensation is discussed alongside intrinsic channel compensation.
Intrinsic compensation mechanisms

**Nature:** Intrinsic compensation for channel distortion describes a process of perceiving target cues relative to other target cues within the same sound (e.g. the same vowel). In particular it appears to describe the process of perceiving targets in relation to particular targets that are reference cues for the colouration caused by the channel (e.g. F0 or F3 may describe the variation to the spectrum caused by variation in the length of the VT). There is a normalisation or process using these reference cues whereby the other targets are heard as if they are shifted back to their canonical location. This process also appears to involve the perception of target ratios rather than absolute values. However, it is not currently clear if intrinsic compensation involves normalisation to reference cues, ratio perception or both. Immediate normalisation to a reference cue and the perception of target ratios appears to involve a similar process but there may be a distinction. If particular targets such as F0 and F3 are more important to normalisation than other targets present in the phoneme, then this implies that there is a normalisation to these particular indicators of the nature of distortion. If all targets are equally prominent then it is more likely that the process can be described as one of relative perception (i.e. the perception of target ratios). Tests like those of Miller (1989) and Syrdal and Gopal (1986) that examine the importance of particular potential reference targets such as F0 or F3 compared to other formants may shed light on the process. The work in this area should start with further examination of the literature on intrinsic normalisation already conducted before an examination of the importance of specific targets. It was noted that a reference cue such as F0 and F3 can only account for variation in the overall length or volume of the VT and does not provide a complete picture of the shape of VT. It is therefore likely that compensation via this mechanism is inaccurate, as not enough information is provided for compensation for the broadband spectrum of the channel compensation. Perception based on ratios may provide a more complete normalisation if multiple ratios are used rather than just one or two, as was suggested by Peterson (1952). This issue should be investigated further.

**Time course:** Like target perception and categorisation this process is *immediate* and reaction time experiments and physiological studies may be necessary to establish time course. Establishing the time course is of interest to establishing the brain mechanisms
behind intrinsic normalisation, but may not be of interest to establishing the scope of the effect. Intrinsic compensation appears to be similar to extrinsic compensation over a short time course (i.e. the enhancement effect and overshoot). It may be useful to further investigate how different intrinsic mechanisms and short onset extrinsic mechanisms are in terms of time course. There may be no clear distinction between the two. For example, listening to prior speech in the form of a transition into a vowel may be enough to shift the perception of the nucleus of a vowel and therefore compensation similar to the enhancement effect occurs, but within a single phoneme.

**Magnitude:** If there is any compensation due to intrinsic compensation more research is needed to investigate the extent to which this occurs. As with other intrinsic mechanisms there is no uncompensated baseline against which to measure perception. Perception can be compared to that expected based on objective measures. As with enhancement tests, shifts in perception after very small prior sound segments (tens of milliseconds) might provide a way of measuring quick acting compensation. The fact that the channel is heard over short listening time courses shows that intrinsic compensation is not complete, but it is possible that this instant mechanism may reduce the effect of the channel considerably compared to the situation where such a mechanism was absent. Some studies appear to show that mixed talker phoneme lists are perceived with higher accuracy than might be expected, suggesting the use of an intrinsic compensation process. However, it is not always clear that these studies have measured perception in regions of overlap. Tests that specifically compare perception in regions of phoneme overlap to non-overlapping regions may be useful for determining the extent to which intrinsic normalisation occurs.

**Mechanisms:** The mechanisms of intrinsic compensation were not discussed in detail here but may involve relative perception, which probably occurs via a cognitive process that has a near instant neural basis (e.g. there may be special populations of neurons tuned to responding to target ratios and these may be able to act quickly to provide compensation). The peripheral neural adaptation mechanisms discussed in this thesis appear involve a too long time course to account for a compensation mechanism that is instant.

**Extrinsic compensation for the VT**
This effect appears to be an example of the spectral compensation effect, so studies leading on from Ladefoged and Broadbent’s (1957) work on extrinsic compensation for the VT are discussed under the heading of the spectral compensation effect.

**The acoustic histories effect**

This effect is likely to be an example the spectral compensation effect so studies leading on from Holt’s (2005, 2006) work are discussed under the heading of the spectral compensation effect.

**The spectral compensation effect**

**Nature:** There are multiple future studies that could be conducted to further determine whether the spectral compensation effect causes channel compensation. The spectral compensation effect was first researched by Watkins (1991) and appears to show an extrinsic mechanism that compensates for the VT and other transmission channels. It appears to provide compensation for the channel via perception that is relative to the LTAS of the channel. It was concluded that the compensation for the VT effect revealed by Ladefoged and Broadbent’s work and Holt’s acoustic histories effect appear to be a manifestation of the spectral compensation effect (Huang and Holt 2012, Watkins and Makin 1994). It should be confirmed that these are the result of the same underlying mechanism. A test of moderation was conducted by Huang and Holt to determine the similarity of the compensation effect with LTAS manipulated precursors (i.e. Watkins’ precursors) and formant range manipulated precursors (i.e. Ladefoged and Broadbent’s precursors). A test of mediation was conducted by Watkins and Makin (1994) to determine whether LTAS manipulated precursors explain the formant range manipulated precursor effects seen in Ladefoged and Broadbent’s VT compensation studies. Further tests using moderation and mediation analysis can be conducted to determine the equivalence of mechanisms. Effects with Watkins’, Ladefoged and Broadbent’s and Holt’s precursors should all be measured in the same test with both mediation and moderation analysis applied. This study should compare compensation with all three types of precursor on the same test sounds and effects should be tested for similarity in nature, time course and magnitude with a variety of different test sounds.
Range: As well as the effects appearing to be of the same nature and size across studies it is claimed that in all cases the effects show compensation for the LTAS of the precursor and that the range of spectra presented is not relevant. This is contrary to Ladefoged’s original suggestion that perception was relative to the spectral range of the talker but Watkins showed the same effect regardless of whether the LTAS of the sentence or the formant range was manipulated. Holt also showed that the range and distribution of the spectra making up the precursors did not have an effect on compensation. Therefore range does not appear to be important. However in Watkins’ work LTAS was not manipulated independently of range, in that study the LTAS probably changed the formant range. Also Holt has noted that it was unexpected that there was no distribution effect, which is seen in compensation in vision. Further work could be undertaken to confirm that compensation is for LTAS and that spectral range and distributional characteristics of precursors have no effect when these factors are manipulated independently. If it is found the LTAS is the only relevant aspect then this supports the conclusion that the mechanism is one that compensates for the channel, as compensation for the time-invariant effect of the channel only appears to require compensation for LTAS. Holt (2006) has examined the effect of sampling density. Further work should be conducted in this area to determine whether the effect requires a sampling of a range prior spectra to building up a picture of the broadband channel response.

Centrality: The spectral compensation mechanism has been shown to be central. This was established with crossaural tests by Watkins, and by showing that it has a time course too long to be peripheral by Holt, but this was not tested by Ladefoged and Broadbent. The best test for centrality is a crossaural test. Future experiments could test Holt’s acoustic histories effect and Ladefoged and Broadbent’s effect with crossaural presentation to confirm that a central effect is seen in those cases.

Direction sensitivity: It was shown in Watkins’ (1991) work that the effect may be direction sensitive. Watkins suggested that this supports the role of this effect in channel compensation as it prevents compensation for one channel being applied to sounds from another. This is the only study to have tested this. Whether this occurs and whether it is advantageous to a channel compensation mechanism requires further research. Direction sensitivity implies a more complex mechanism than one relying
on neural adaptation or memory—neither of which themselves are inherently direction sensitive. Whether a direction sensitive effect can compensate for colouration carried by room reflections should also be examined.

**Time course:** The time course of this effect was suggested to be long, requiring a sentence to build up, with its offset being 10 seconds or more. However, the time course is not well established. It should be confirmed that the onset of this effect is indeed longer than enhancement to firmly distinguish the mechanisms. There is evidence that the onset may begin with a phoneme (a short phoneme brought about a backwards acting central effect in Watkins and Makin’s (1996b) work and there was evidence for a central effect with overshoot studies, which only use a phoneme length precursor). The nature of the spectral compensation effect is likely to change as LTAS changes due to the prior segment getting longer. This should be confirmed. It was also suggested that the effect may become larger with increased precursor length. These features appear to be in accordance with what might be expected from a mechanism that compensates for the channel. It is expected that mechanisms that compensate for channel would begin to onset immediately, applying inaccurate compensation at first but more accurate, and perhaps stronger, compensation as a picture of the channel is built up. These features should be confirmed. The offset length should be further researched. It should be confirmed that this is longer than enhancement to firmly distinguish this mechanism from enhancement. Evidence suggests that it may be over 10 seconds. The role of longer onset in building up a channel picture is clear but whether longer offset is also beneficial could be investigated. For example, it could be determined whether this is necessary, as Watkins suggests, to prevent gaps in compensation where there are gaps in running programme items. It may simply be a side effect of a long onset process.

As well as confirming that the time course is different from enhancement, establishing the time course with acoustic histories precursors, Watkins style precursors and Ladefoged and Broadbent style precursors is a means by which the effects elicited can be compared. Time course differences may suggest different mechanisms are at work in these studies. Once the time window for the effect is established an auditory model of compensation via the spectral compensation effect should be made and tested against perceptual data.
Magnitude: The magnitude of this effect was not well described. The measures of compensation vary between studies. All measures of the spectral compensation effect measure shifts in the categorisation of test vowels but because different stimuli and conditions are used it is not possible to compare shift sizes. A single test should be conducted whereby shifts to the same test sound are observed with the different precursor types. Further work that shows that the magnitude of the spectral compensation effect is similar across precursor types would demonstrate that the effect is caused by the same mechanism in all cases. Tests to determine the size of the spectral compensation effect compared to enhancement should be conducted to further distinguish these mechanisms. Tests to determine whether the effect becomes larger as the picture of the channel is built up should be conducted. A channel compensation mechanism may be expected to produce larger effects as it becomes more accurate.

More generally meaningful ways of quantifying compensation should be considered. The size of shifts in vowel test sounds cannot currently be compared to the shifts that occur with non-speech test sounds. It may be preferable to measure channel perception based on dB reduction (or reduction in Phons). Whether or not existing findings can be translated into such a measure should be considered. Further, studies that measure source identification accuracy measure channel perception indirectly; future work should consider direct ways of measuring channel perception, such as asking listeners how their perception changes when the same source is presented through different channels (as done in the real-world compensation studies presented in Chapter 2).

The extent to which the magnitude of compensation via the spectral compensation effect is sufficient to ameliorate the effects of the VT and overlap caused by VT variation should be investigated further. The extent to which the mechanism compensates for colouration caused by real transmission channels such as loudspeakers and rooms, which cause milder colouration compared to the inverse-vowel channels seen in the speech perception studies, should be examined.

Mechanisms: The mechanism has been shown to be central but as mentioned above, further tests could be run to confirm this. The mechanism originally suggested to be behind extrinsic compensation for the talker was relative perception (Ladefoged
and Broadbent 1957). This implies that a cognitive process is involved. Watkins put forward a memory explanation for the effect. It was argued in this thesis that memory is unlikely to explain compensation for the channel but work to determine the role of memory in this effect and in compensation for the spectrum generally is described in Section 7.2.3.

Watkins suggested that spectral variation may be necessary for the effect to occur. Further work is needed to establish the importance of spectral variation. The requirement of spectral temporal variation appears to be incompatible with simple adaptation effects, as this type of adaptation is thought to occur more readily for static sounds. This author suggests that spectral variation in the precursor might allow for the sampling of the whole channel response, which might be necessary to build up a picture of LTAS. However, as suggested by Watkins, spectro-temporal variation within the precursor might also result in perceptual ungrouping of the source from the channel. This might be necessary for compensation. Holt suggested a central neural adaptation mechanism that tracks reliability and enhances the perception of change might cause the spectral compensation effect (stimulus specific adaptation). The backwards acting nature of the effect should be investigated further. This suggests that an adaptation process is not involved because adaptation is forward acting. It appears that only a complex mechanism would explain any backwards acting effect. It should be confirmed that this occurs, as this has only been tested in one paper and may have been an artefact of the experimental process (no backwards acting effect was found for enhancement in Summerfield’s work). If the mechanisms can be shown to be different for the compensation seen in Ladefoged, Watkins or Holt’s studies then this will show that spectral compensation is not behind compensation in all of these studies.

**Non-speech perception:** It is unlikely that the effect is a result of a speech mode of listening and due to a process that occurs with speech but not other sounds. The work in this thesis confirmed a central effect in a non-speech test. However, further work could be done to confirm that it was the spectral compensation effect that caused the effect seen in that study. For example, the time course of the effect could be tested and the fact that it is applying compensation for the LTAS of the whole musical segment rather than the immediately prior spectrum should be confirmed.
7.2 FUTURE WORK

The overshoot effect

This effect is likely to be an example of the enhancement effect so suggested studies leading on from overshoot work are discussed under the heading of the enhancement effect.

The enhancement effect

Nature: Enhancement appears to cause compensation for the LTAS of the immediately prior phoneme or phonemes-worth of sound. This enhances spectral change between phonemes and provides overshoot. It has also been hypothesised to be involved in channel compensation. The role of enhancement in creating a reduction in perceived channel effect requires further testing. Watkins observed that the effect is small, would break with gaps in running programme items and is not direction sensitive so does not have the necessary features for a channel compensation mechanism. It has been suggested in this thesis that the time course of this effect is only long enough to cause compensation for the prior sound not the channel. The nature of compensation applied cannot provide channel compensation because adaptation is in respect of the prior phoneme, rather than the spectrum of whole channel. This is because the whole channel is not excited with every phoneme therefore compensation applied when hearing a current phoneme can only be for the spectrum of the prior phoneme plus the portion of the channel that is excited with that phoneme, not for the channel as it is present when hearing the current phoneme. This process would in fact involve different spectral regions of the channel being enhanced with the presentation of different phonemes. Only a longer time course adaptation can remove the whole channel response with each phoneme. A computer simulation of the adaptation process should be made to confirm this. Further work to establish the magnitude and time course of the enhancement effect is necessary before such a model can be produced. This model should also aim to demonstrate the role of enhancement in overshoot as well as channel compensation.

Further tests could also be run to confirm that the shifts seen with enhancement are in relation to the immediately prior spectrum but not LTAS of the full longer segment. For example a test that shows a monaural shift is different in spectrum to a cross-aural
shift with long precursors may show this difference as the monaural shift will only be relative to immediately prior spectrum, whereas the cross aural will only be relative to LTAS of the channel.

**Time course:** There are varying estimates of the time course of the enhancement and overshoot effects. These are broadly in line with a mechanism that enhances spectral change between phonemes and produces overshoot; however, further work is necessary to confirm this. It is necessary to further establish the time course of enhancement using all three variations on the enhancement effect paradigm described in the literature (Summerfield’s, Watkins’ and the Overshoot paradigm). This will provide more precise information on the time course of effects and will confirm the similarity of overshoot and enhancement in terms of time course. Onset with a phoneme and offset within a couple of hundred milliseconds should be confirmed if the hypothesis that this effect results in overshoot is to be supported. If the effect takes longer to onset than the duration of a typical phoneme compensation will be more for the channel rather than the phoneme spectrum. Likewise if the offset is too long then compensation becomes for the channel rather than the phoneme. It is not currently clear that the time course is a good match to spectral change between phonemes or that the effect of enhancing a current phoneme’s new energy relative to that of the prior will successfully bring about an enhancement of change between phonemes. An auditory model that uses the time course of the effect should establish the effect of enhancement on perception in running speech to determine whether this aids perception of spectral change between phonemes and whether it contributes to overshoot as well as channel compensation. The overlap between reduced masking thresholds with prior presentation of the masker and the enhancement effect should be further investigated. It is possible that a different process is involved in each.

**Magnitude:** The magnitude of this effect is largely unknown. It has been difficult to measure magnitude in a meaningful way. Some enhancement studies show that the process entirely removes the perception of the channel (e.g. no perception of the context vowel in DSV studies) while others appear to show a small reduction of the channel and others a small boost for the spectral change, but no reduction of the channel. It is not clear whether the channel effect is stable but change enhanced, or whether the channel effect is specifically reduced alongside the enhancement of change. Further, the extent
of any enhancement/reduction is not known. The reduction in the channel has been measured directly by the extent to which the channel is perceived and indirectly by the boost in identification accuracy of the source. The magnitude has been measured in dB, change in percentage correct identification of speech sounds and shifts in tendency to categorise a vowel as one category or another. It is currently extremely difficult to make comparison of effect sizes across studies. Future work should aim to determine a meaningful measure of the magnitude of compensation via enhancement that can be used in all studies. Ideally, perceived reduction in channel perception measured in dB (or Phons) would be the most direct way of measuring channel compensation. But the extent to which this improves the identification of sounds within the channel and provides a channel-free perception of those sounds may be more relevant where the effect of channel reduction on timbral constancy is of interest. It should also be established that enhancement has the magnitude necessary to compensate adequately for the effect of undershoot and, if it is considered a channel compensation mechanism, the magnitude necessary to compensate for overlap caused by the VT or colouration caused by other channels.

**Mechanisms:** The mechanisms behind enhancement are thought to be low-level adaptation processes. This should be confirmed via further experiments to show that the time course of effects matches the time course of peripheral adaptation. More physiological research should be examined for evidence of a neural basis in auditory periphery or elsewhere. Other quick acting central adaptation processes may be considered. More research into the role of adaptation via the MOC feedback process as an explanation of the enhancement effect should be conducted, as this is shown to have a short time course and reduces the perception of additive energy due to background noise and room reflections. This mechanism may also reduce the static additive/subtractive energy to a source that is due to spectral filtering. In this review evidence of a neural basis of the enhancement effect was searched for but this did not extend to the neural bases for decreased masking with prior presentation of maskers or the neural basis for the overshoot effect. Research in these areas is likely to provide information on the neural basis for enhancement effect.

Summerfield et al. (1984) suggested that categorisation, attention, and perceptual grouping may be behind enhancement. But it was concluded by Summerfield that
these are probably beneficiaries of the enhancement effect rather than causes. However, further work should be done to investigate this claim. Mechanisms underlying perceptual grouping may have a role in causing the enhancement effect. Further work should be undertaken to determine the similarity of overshoot and enhancement in terms of mechanisms. If the two effects are shown to have different mechanisms then they cannot be both instances of the enhancement effect, as has been suggested.

**Non-speech:** Enhancement appears to be a general auditory processes. A number of studies point to this and Experiments 4 and 5 in this thesis show a monaural effect in an all non-speech test with noise and music. However, further work must be done to confirm that the monaural effect seen in Experiments 4 and 5 is monaural and not a central direction sensitive effect.

### 7.2.2 Further tests to determine real-world compensation mechanisms

This section describes future work to establish the *actual* role of the potential mechanisms discussed above in the time-gap sensitive component of compensation (Step 2a). In Section 5.4 it was concluded that the enhancement effect and the spectral compensation effect might explain the time-gap sensitive component of compensation by enhancing timbral differences when stimuli are heard side-by-side, but not with time gaps between them (i.e. they enhance differences when stimuli are compared directly but not indirectly). To test this, the real-world paradigm can be used and the unique features of each effect can be used to exclude each and therefore test for their role in compensation. The features of the spectral compensation effect that may be used to distinguish this from other effects are its central origin, longer time course, and the fact that it compensates for the LTAS of a longer prior segment, rather than the immediately prior segment. Enhancement can be distinguished by the fact that it is monaural and is extrinsic with a short time course, applying compensation for the immediately prior phoneme or phonemes-worth of spectrum.

To test for the role of the enhancement effect in the time-gap sensitive component of compensation its monaural rather than central nature can be used to exclude it. A
crossaural test using the real-world paradigm can be run whereby the channels being compared are never presented side-by-side to the same ear, but always to the ear that is contralateral. Therefore, all comparisons are made between stimuli presented to different ears. This crossaural presentation excludes any monaural enhancement of timbral differences that occurs between adjacent channels via enhancement. However, this does not prevent the enhancement of timbral differences occurring due to the spectral compensation effect which occurs crossaurally. If the enhancement of timbral differences between channels via the enhancement effect is a cause of the compensation seen in real-world tests then it is expected that compensation would be reduced compared to where the enhancement effect can enhance timbral differences (specifically the differences between direct comparisons will be reduced, as some of what causes larger differences when stimuli are heard side-by-side has been removed). If the compensation in real-world tests is fully explained by the enhancement of adjacent stimuli caused by the enhancement effect then real-world compensation will not occur at all—direct comparisons will be as small as indirect comparisons.

However, the lack of enhancement of channel differences via the enhancement effect when directly compared is not expected to fully explain the compensation seen in real-world tests. It is expected that this process may only partially contribute to compensation. A reduction in compensation may be expected, with larger timbral differences still being reported for direct comparisons compared to indirect (perhaps due to enhancement of side-by-side differences via the spectral compensation effect, the fatiguing that comes about with longer listening with indirect comparisons, or other compensation factors). The reduction of effect in the direct comparison condition compared to the situation where enhancement is present is a measure of the role of the enhancement effect in compensation.

It is expected that if it does contribute to real-world compensation the enhancement effect will contribute mainly or entirely to the time-gap sensitive component of compensation. This is because differences between adjacent stimuli are no longer enhanced with indirect comparisons. It appears to contribute little or not at all to the non-time-gap sensitive component (i.e it does not appear to contribute to compensation that occurs with longer listening present in the indirect comparisons, this is because enhancement is at maximum with both direct and indirect comparisons.
unlike the spectral compensation effect, which may show more compensation with indirect listening as it takes longer to reach maximum).

The extent to which this mechanism contributes to the time-gap sensitive portion of compensation specifically can be tested by examining the extent compensation is reduced with the elimination of the time gap between indirect comparisons after enhancement via the enhancement effect is excluded. Under such circumstances as the enhancement effect is no longer working to enhance either direct and indirect comparisons, if enhancement is solely responsible for the time-gap sensitive component of compensation removing the time gap between indirect comparisons is not expected to increase timbral differences (i.e. no time-gap sensitive component will be seen). If another effect is also contributing to the time-gap sensitive component then removing the time gap is still expected to show increased differences between indirect comparisons compared to where there is a time gap. Any time-gap effect remaining cannot be due to enhancement and would show another time-gap sensitive compensation mechanism at work, which must be central in nature. Such a mechanism would likely be the spectral compensation effect, as this is the only central time-gap sensitive compensation mechanism revealed by the literature review (though see Section 7.2.3 for other possible explanations). The difference between the effect of time gap with and without enhancement excluded can be used to determine the enhancement effect’s contribution to the time-gap sensitive component of compensation. Because the enhancement of timbral differences via the enhancement effect is small and short lasting, it is expected that it might only explain a small proportion of the time-gap sensitive component of real-world compensation. It is expected that most of this will be due to other compensation effects, namely the spectral compensation effect. However, it is preferable to measure the contribution of the enhancement effect before other effects as this is easy to measure and may allow for a simple determination of one potential source of real-world compensation.

Before this enhancement test is run it is advisable to carry out the recommendations made in the conclusion to Experiment 3. In Experiment 3 the time-gap sensitive component of compensation was measured and it was only shown to occur for the room factor. It was noted that it is possible that ceiling effects may have prevented this from being shown for the loudspeaker factor. It is preferable to show that compensation
for both channels being tested has a time-gap sensitive component before proceeding with examining the mechanisms behind this source of compensation further. A replication of Experiment 3 should be made whereby the time-gap sensitive component of compensation is measured in a situation where ceiling effects are prevented.

Further, the effect of time gap was small. Methods should be used to ensure that future investigations that test for a time-gap sensitive component of compensation are not subject to error, which may prevent small effects being regarded as statistically significant. If the time-gap sensitive component is small and the enhancement effect only explains a portion of this, a change in the time-gap sensitive component with the elimination of the enhancement effect may not be statistically significant unless statistical uncertainty is made very small, which is best achieved by obtaining many more data points than used in previous experiments. It is also noted that the overall error in judgements may increase in tests with crossaural comparisons of the stimuli, rather than binaural comparisons as this is a more difficult task. Therefore, steps to reduce error are even more important.

It is also legitimate to attempt to increase the magnitude of the enhancement effect for the purpose of this experiment. This would result in a test that determines the importance of the enhancement effect to the time-gap sensitive component of compensation under conditions most conducive to the enhancement effect contributing to this. The role of the enhancement effect could be boosted by ensuring that the channels being compared are not very different, as the enhancement effect appears to be strongest where timbral differences between the compared stimuli are small. Too large channel differences may result in a categorisation of channel differences that may prevent enhancement from boosting perceived timbral differences further (i.e. ceiling effects may occur). However, as was noted in Experiment 1, it is also important that the timbral differences between channels are large enough to be significant when making direct comparisons so as to prevent floor effects and allow for significant compensation to occur.

Enhancing the role of the enhancement effect in the time-gap sensitive component of compensation may mean that the role of the spectral compensation effect in this appears smaller. To properly establish the role of enhancement and the spectral
compensation effect in the time-gap sensitive component of compensation tests should be run where the strength of the enhancement has been both maximised and minimised.

While it has been concluded in this section that any time-gap sensitive component of compensation remaining must be caused by the spectral compensation effect, there remains a possibility that other central mechanisms not currently regarded as established compensation mechanisms (and therefore were not included in the literature review) may explain the time-gap sensitive component of real-world compensation. One example of such a mechanism is timbral memory. The potential role of this and other mechanisms is discussed in Section (7.2.3). It is possible to run an equivalent test for the spectral compensation effect to determine the extent to which this explains the remaining time-gap sensitive component of real-world compensation directly. This may be tested by excluding this effect, as is done to test for enhancement. However, it is more difficult to exclude this mechanism. For this reason it may preferable to rule out all alternative mechanisms as explanations of any time-gap sensitive compensation that remains after enhancement is excluded, rather than to test for the role of the spectral compensation effect directly. Therefore tests for other possible explanations are discussed in the next section, Section 7.2.3.

7.2.3 Other explanations of the time-gap sensitive component of compensation

There may be causes of the time-gap sensitive component of compensation other than the enhancement and spectral compensation effects. Firstly, there may be a short onset central effect, as was suggested by the work of Holt and Lotto (2002). A test of the time course of the central effect seen in Experiments 4 and 5 would be useful to determine this. It is unlikely that such an effect exists as the effects seen in Holt’s study may be explained by the spectral compensation effect (see Section 5.3.1), but if it does, then this may contribute to the time-gap sensitive competent of compensation. Other than this, further explanations of the time-gap sensitive component are speculative. Any other mechanisms that might explain the time-gap sensitive compensation are not currently regarded as compensation mechanisms. An
aim of this thesis was to investigate the role of established mechanisms of channel compensation (i.e. enhancement and the spectral compensation effect) in real-world compensation, rather than more speculative mechanisms. Ideally, the role of the established compensation mechanisms would be tested directly using the real-world paradigm before speculative mechanisms are tested (this aim was set out in Section 4.2 in Chapter 4). However, as was stated in Section 7.2.2 above, when examining the time-gap sensitive component of real-world compensation it is possible to exclude the enhancement effect but more difficult to exclude the spectral compensation effect and test for this directly. Therefore, it is preferable to rule out alternative speculative mechanisms, leaving the spectral compensation effect the only possible explanation. One such mechanism that may explain the time-gap sensitive component of real-world compensation is timbral memory. This mechanism was first suggested as a compensation mechanism by Olive et al. (1995) who hypothesised that real-world compensation (compensation when stimuli in his study were indirectly rather than directly compared) was due to memory loss with the time gaps between stimuli. However, it has been concluded by this author that memory is unlikely to be the cause of this real-world compensation. This is because it is not an established compensation mechanism—memory is not currently well known for its role in compensation. More importantly, it does not appear to have the required nature to explain compensation. This nature is discussed in more detail here and further work to confirm this conclusion is discussed. As well as timbral memory directly causing real-world compensation another way that it may be involved is through the spectral compensation effect: Watkins (1991) observed that memory may be a cause of the spectral compensation effect. Whether or not timbral memory explains real-world compensation, either directly or as the mechanism behind the spectral compensation effect, is discussed below.

Timbral memory appears to have the appropriate time course to explain the time gap sensitive component of real-world compensation; when listening is done over a shorter time course (as with direct comparisons) memory for timbral differences between stimuli is good and timbral differences may be more apparent, when listening is over a longer time course (as with indirect comparisons) memory may be poor and timbral differences may be less apparent. Any timbral differences between sounds caused by channel
differences may be less apparent with poor memory. Specifically, the listener may perceive smaller differences between stimuli due to *memory decay*, which occurs with time between listening (Cowan 1984).

However, memory decay does not appear to have the required nature to explain the smaller timbral differences between channels seen with indirect comparisons. As a speculative mechanism, timbral memory was not covered in the literature review but a brief review of the memory literature reveals that memory loss through decay might not result in the contraction of perceived differences between stimuli seen in real-world tests and therefore it would not be expected to bring about real-world compensation. Instead, timbral memory is more commonly associated with an increase in uncertainty as to the nature of a difference between stimuli. Therefore, memory loss is likely to be a source of uncertainty or *noise* in judgements rather than a cause of increased perceived similarity. For example, when rating timbral preference using a continuous rating scale (as was done in the real-world tests described in this thesis) memory loss would be more likely be reflected by increased variation in the exact rating given to stimuli across repeated ratings but the mean ratings would not be expected to contract. The contraction of mean ratings with indirect comparisons appears to require a compensation mechanism that has something other than memory as its basis—adaptation or relative perception appear to be better explanations.

However, the lack of current evidence showing a contraction in ratings due to memory loss does not necessarily mean that memory loss does not cause such a contraction. It is possible that memory loss is a cause of adaptation/relative perception, which *do* cause such contractions in mean ratings. Further work should be conducted to confirm whether memory loss results in adaptation/relative perception as well as/instead of an increase in noise in judgements. Work which clarifies this issue has not been found because, in previous literature, the effect of memory loss has been measured using binary outcome measures. Common paradigms used to test memory are recall paradigms (accurately vs not accurately remembered) and recognition paradigms (percentage of items accurately discriminated as being old compared to new) with various time gaps between stimuli presentations. Performance accuracy is therefore measured using binary decisions and the percentage correctly remembered is calculated. Decreased percentage of item correctly remembered is used to indicate memory loss.
In a timbral memory test of this style, timbral memory could be measured by asking whether the listener correctly identifies a new stimulus as containing more or less HF than a standard, with time between the pair, and the percentage correct responses measured. However, the percentage correct measure would confound the effect of decreased certainty in judgements that may occur with memory loss and decreased perception of timbral difference that may also occur with memory loss. For example, when comparing sound ‘A’ to a previously heard ‘X’ for more HF content, it is not certain whether a decrease in accuracy in judgements due to memory loss reflects the listener perceiving ‘A’ and ‘X’ to have a constant level of HF content on average (for a given difference in HF between ‘A’ and ‘X’) but with increased variation in perception of ‘X’ over repeated ratings (due a noisy perception of this), leading to more responses to ‘A’ in the incorrect region or, whether the variance in perception remains constant, but the average perception of ‘X’ moves towards the current sound, ‘A’ with increasing time gap, so the listener remembers the ‘X’ to be less extremely different to the ‘A’. When using such a binary measurement method in the latter case, like the former, more responses across repeated ratings would fall into the incorrect region because of this shift in locus of perception, causing a reduction in performance accuracy. Both scenarios result in a lower percentage correct for the binary measure but only one shows compensation. However, if ratings were made using a continuous scale the mean responses would simply appear closer together if compensation due to time-gaps (memory loss) could be seen alongside/instead of any increase in variance (increased uncertainty in perception).

Memory tests that use traditional percentage correct measures do not aim to separate these two potential explanations of this decrease in accuracy. It is suggested that only a contraction of stimuli differences shows compensation, as this shows perception that is relative to the context and a (neural or cognitive) adaptation mechanism. Tests should be undertaken that can show which of these processes occurs when judging timbral differences with time gaps. This can be done by using a continuous measurement scale (such as that used in the real-world experiments described in this thesis) and measuring contraction versus increased error when stimuli are judged with different time gaps between them. As the time gap between comparisons is lengthened the enhancement effect, spectral compensation effect and short-term timbral memory are all expected to
cease working. All of these effects may be expected to cause an expansion in timbral
differences when they are working and a contraction when there is a long enough
gap between stimuli for these effects to offset. The known offset time of these effects
can be used determine which processes are causing any contraction in ratings seen
(approximately 500 ms is the current estimate of the offset of the enhancement effect,
10 s is the estimate for the spectral compensation effect, and short-term timbral memory
lasts approximately 30 s). It is therefore, expected that there will be a contraction in
ratings with over 500 ms time gaps between stimuli (due to the absence of enhancement)
and a further contraction with over 10 s time gaps due to absence of the spectral
compensation effect. Any further moving apart of the stimuli in time is predicted to
cause memory loss only. Any further contraction of ratings is seen after this time,
suggests that timbral memory also causes a relative perception/adaptation that can
explain real-world compensation. However, if memory loss is not a compensation effect
then only uncertainty (noise) is expected to increase after this time. Uncertainty
would be expected to be low up to 30 s where short-term timbral memory is thought
to be good but higher after this time when short-term memory decays or offsets. If
any contraction of ratings only matches the time course of enhancement and spectral
compensation effects, but not short-term memory (or long-term memory), then it may
be concluded that memory does not cause compensation. Results could be compared
with binary performance accuracy measures. A drop-off in performance accuracy would
be expected where enhancement and spectral compensation cease to work. A further
reduction in accuracy is expected where short-term timbral memory also expires.

Problems that may prevent the measurement of these effects separately are: a) the
time course of the spectral compensation effect may be longer than current estimates
have shown—it may be as long as short-term memory. Therefore, it would not be
clear whether a contraction of ratings is due to the spectral compensation effect or
memory. It is important to clarify that the offset time of the spectral compensation
effect is different to that of short-term memory before this test is conducted; b) it is also
possible that long-term memory will occur to prevent an increase in uncertainty with
longer than 30 s time gaps, thereby preventing measures of noise being used to establish
a difference between memory and spectral compensation mechanisms. However, stimuli
that are less susceptible to storage in long-term memory may be used to prevent this.
After conducting this test more will be known about the nature of timbral memory loss. If the term *memory loss* simply describes a process by which accuracy in perceiving timbral (or other) differences becomes reduced with time gaps when using a binary measure, then *memory loss* may be regarded a compensation process as some of this decrease in accuracy of perception appears to occur due to lack of the enhancement of timbral differences via an adaptation processes (e.g. the enhancement effect and spectral compensation) and/or relative perception. But if the term *timbral memory* describes a process by which performance accuracy reduces due to an increase in uncertainty with time between stimuli, then memory is not a compensation process. It is possible that memory is a separate process to compensation, with memory modulating certainty and another process (e.g. enhancement/spectral compensation mechanism) modulating compensation. Both change with time and longer time between stimuli results in the listener appearing to be less sensitive to timbral differences (when measured in a binary manner). Memory’s role in genuinely reducing perceived timbral differences, rather than increasing noise, between sounds may depend on the extent to which we judge timbre in a binary manner (e.g. brighter versus darker; more pleasant, less pleasant) in everyday listening.

A related question that remains is whether memory rather than neural adaptation explains the enhancement and spectral compensation mechanisms, as suggested by Watkins (1991). More specifically, whether shifts that are thought to be caused by enhancement and spectral compensation (thought to have a neural adaptation basis) can arise from *relative perception* that has a cognitive basis. If this is the case, then the cognitive basis might be memory. The compensation seen in Experiments 4 and 5 show relative perception may occur via enhancement (as hypothesised by Summerfield et al.) and the spectral compensation effect but, as was discussed in Section 6.1, there may also be other means of causing relative perception. Relative perception is the process by which stimuli are compared to a prior stimulus, or prior stimuli, and perceived in relation to this (a mid-frequency timbre sound is heard to contain more HF than expected when heard after a prior sound with reduced HF content, compared to when heard in isolation). Memory appears to be involved in this relative perception. Without memory for the prior sound each sound would be judged in isolation, rather than with regard to its position relative to prior sounds. In this sense, relative perception
is caused by memory and memory might be said to cause compensation seen with indirect comparisons. Specifically, without memory (e.g. with time-gaps) there may be a tendency to hear the new (isolated stimuli) as being the perceptual mean or midpoint. But as long as there is short-term memory of a prior sound (which would have been previously assigned the perceptual midpoint), the new sound will be heard relative to this, rather than being heard to be as the midpoint itself (e.g. if the prior sound had low HF content and the new sound had high HF content, the new sound may be heard as containing even more HF (very high HF content) than it would do if heard in isolation where it would be heard as having a middle spectrum (a mid HF sound) or heard in its absolute position (a high HF content sound)). It is unlikely that a cognitive memory process causes the enhancement effect, due to its non-central nature but it is possible that memory is the cause of the spectral compensation effect (which is central), explaining the shifts in perception that result from this.

Memory may be considered instrumental in causing relative perception and may therefore be considered a process which causes compensation. However, it may be that memory is involved in relative perception in some respect, but it is not clear that memory causes relative perception. Memory may just facilitate this or mediate this. There appears to be something additional necessary to dictate that the listener will listen with reference to short-term memory, but will not listen to the sound as if in isolation and/or with reference to long-term memory. Therefore memory does not appear to completely explain why relative perception occurs. It appears that a neural adaptation process accounts for our tendency to listen with reference to short-term memory well. Specifically, the time course of saturation of responses in neural adaptation explains why perception is relative to the short-term prior context and not in isolation or in relation to the longer-term context. Further, it is not clear how memory, after storing a train of previous sounds, builds up a picture of an LTAS and applies this in compensation. However, neural adaptation processes appear to easily explain this. The LTAS is built up by the ongoing fatiguing of neural responses and the reduced responsiveness in the face of fatigue causes compensation for this LTAS. It therefore appears that neural adaptation explains both compensation and short-term timbral memory.

In further support of this hypothesis, there has been recent research that has described
short-term timbral memory as a neural adaptation process. This research indicates that it may be that the spectral compensation effect (as a centrally produced neural adaptation based process) is the very process that underlies timbral memory, rather than the other way around (McKeown and Wellsted 2009). Further research should be conducted to confirm that the spectral compensation effect has a central neural adaptation basis and also determine whether timbral memory has the same basis to support this hypothesis.

Against the argument for memory as the cause of the shifts seen in Experiments 4 and 5 is the fact that centrally produced shifts indicative of compensation have not been shown to occur with noise (i.e. there is no centrally produced relative perception when listening to noise). Memory for timbre might be expected to act equally when listening to music and noises so this calls into question memory as a cause of the shift seen in compensation studies. Memory for a sensation (and memory more generally) is currently thought to be a central process which does not depend on the ear, eye, hand to which stimuli are presented. So this finding brings about the unexpected implication that timbral memory does not cause relative perception when listening to noise and, further, that memory does not work when listening to noise. It appears from prior research that the central compensation mechanism requires spectral variation in the sound. Timbral memory is not currently known to require the timbres being remembered to be spectrally varied. However, some evidence exists that memory may be weaker for static noise stimuli (Pike et al. 2014) and timbral memory is known to be weak for static pitch stimuli (only listeners with perfect pitch have good pitch memory). So timbral memory may still be the central mechanism that can explain shifts with spectral varied stimuli but not static stimuli.

Finally, regardless of whether memory causes relative perception there is another way that it might be involved in the shifts towards the mean seen with indirect comparisons in real-world tests. Memory may be involved its influence on the listeners perceptual dynamic range. This explanation assumes that memory only causes an increase/decrease in certainty. Memory may be responsible for the contraction in ratings seen with indirect comparisons via a decrease in certainty that results in memory loss due to the time gaps present. With indirect comparisons the listener may be aware that there is more uncertainty and more variation in the perceptions. They may feel
that larger space on the rating scale is needed to give themselves room on the rating scale to account for this increased variation. In which case they may shift the mean response for each stimulus towards the midpoint of the rating scale. This might occur because they are dealing with a limited rating scale in the experiment (this can be tested by expanding the given scale). But it may also be an effect of dealing with their own limited perceptual range: the listener must make perceptual room for the greater space required on their perceptual rating scale when perceiving stimuli under conditions of uncertainty this may push their mean perceptions towards the midpoint. In this case it can be concluded that memory via increasing uncertainty causes the contraction of ratings due to a limited perceptual range. This effect may depend on whether binary perceptual judgements are made (more versus less preferred, brighter or darker).

It can be concluded that the most likely time-gap sensitive cause of the contraction of ratings with indirect comparisons are the enhancement and spectral compensation effects (with an neural adaptation basis). These processes have the nature to cause shifts that show relative perception. The fatiguing in neural response necessarily results in these shifts. Timbral memory may also be involved in relative perception. If every sound is heard relative to the immediately prior sound (which has become the perceptual midpoint) the same shifts would occur. However, it is not clear why listening would be relative to the short-term context rather than in isolation or relative to the longer-term context (i.e. relative to sounds held in long-term memory). There does not appear to be anything about timbral memory per se that necessitates perception relative to short-term memory. Only a memory that has the above mentioned neural adaptation as it basis does this. Therefore, perceptual shifts that show relative perception and compensation for the prior context appear to be caused by neural adaptation mechanisms rather than memory. Research that shows that an adaptation process like that which appears to be involved in the spectral compensation effect is behind timbral memory may indicate that enhancement/spectral compensation mechanisms are the basis of timbral memory. It is also concluded that the simplest way in which timbral memory may cause the the contraction in ratings seen with indirect comparisons is via it causing increased noise in perceptual judgements and listeners making more room to account for this when using the rating scale—by
shifting their mean perceptions for the varied stimuli towards the midpoint. As the most simple explanation, this should be examined before further work to determine the role of memory in the time-gap sensitive component of real-world compensation is carried out.

7.3 Thesis conclusion

The thesis set out to determine the extent to which compensation for the spectral effects of transmission channels (loudspeakers and rooms) occurs in real-world listening and the psychological/physiological mechanisms involved. The research questions set out in Chapter 1 were as follows:

‘How could perceptual compensation for spectral distortion caused by the transmission channel affect the perception of loudspeakers and rooms in real-world listening?’

This question was divided into 3 sub questions. Research Question 1a asked:

‘In what way can compensation for spectral distortion caused by the transmission channel affect the perception of loudspeakers and listener rooms?’ This question aimed to determine the nature and extent of the effect that compensation may have on perceptions of sounds heard through loudspeakers in rooms.

Research Question 1b asked:

‘By what means can compensation affect the perception of loudspeaker and rooms?’

This question aimed to determine some of the potential mechanisms involved in compensation when listening to music and speech through channels in the real world.

Research Question 1c, which would not be directly investigated as part of this thesis asked:

‘What are the actual mechanisms of real-world compensation for loudspeakers and rooms?’ This question was discussed in Chapter 4 and in this chapter. Answers to this question have not yet been provided, but a test to further answer this question was described in Section 7.2.2 of this chapter.
The answers to Research Questions 1a and 1b were discussed in brief at the end of each chapter. The answers to these research questions are collated here to show the extent to which each research question has been addressed.

**Chapter 1** set the research context and defined terms. The research questions were set out. The following was demonstrated: the importance of timbre perception to auditory object recognition; importance of spectrum to timbre; the potential of the channel to detrimentally affect recognition through distorting a sound’s spectrum and; that timbral constancy may occur to combat this distortion allowing listeners to hear sounds in a channel-free context.

Research Questions 1a and 1b were set out here but not directly answered. However, in relation to Question 1a it was shown that researchers have hypothesised that *timbral constancy* occurs in perception. This implies that there may be compensation for channel distortion. In relation to Question 1b it was noted that timbre is processed at multiple levels in the hearing system so compensation could originate at a number of points along the stimulus-to-perception chain, e.g. peripherally or centrally. There was a suggestion that real-world listening may bring about increased compensation compared to laboratory listening and there may be something about real-world listening that causes compensation.

**Chapter 2** aimed to report evidence of compensation for loudspeakers and rooms from audio engineering studies that have measured the effects of these channels. It was noted that compensation was observed in some circumstances. An instant compensation mechanism may exist but this does not offer complete or large compensation. It was apparent that compensation was greatest over a longer time course of listening. This longer time course of listening is representative of listening in the real world but not in the laboratory. Therefore, additional compensation that occurs in real-world listening appears to occur and may due to longer listening periods. Only one study had appropriately measured compensation that occurs in real-world listening, comparing compensation over a longer time course to that of laboratory scenarios where listening time is short. Compensation was shown but needed to be confirmed.

This chapter partially answered Question 1a by showing limited evidence of real-world
compensation for loudspeakers and rooms. Compensation was seen in a real-world listening scenario with a longer time course compared to a laboratory scenario. This compensation resulted in a diminished perception of the channel. It was noted that further work is necessary to confirm that this and determine its extent. To further measure real-world compensation the effects should be replicated using the same test paradigm.

The work presented did not primarily aim to answer Question 1b but it was hypothesised that real-world compensation may be due to time gaps between hearing different channels in real-world listening and that memory may cause this compensation. It was also hypothesised that the longer continuous listening to channels in real-world listening may cause a timbral fatigue similar to loudness adaptation and this may bring about real-world compensation. However, timbral memory and timbral fatigue are speculative hypotheses. There may be known specific mechanisms of channel compensation that can explain the compensation experienced in real-world listening. A binaural rejection mechanism may occur but does not explain the extra compensation seen in real-world listening compared to laboratory listening.

Chapter 3 aimed to further confirm real-world compensation and measure its extent. One experiment provided some evidence of compensation but this was not significant. A second experiment used conditions more conducive to compensation and showed significant and moderate to large compensation for loudspeakers and rooms. Experiment 2 also provided preliminary analysis to determine whether a mechanism that requires longer continuous listening to the channel causes compensation. It was shown that a mechanism that requires continuous listening between 25 s and 4 minutes is not likely to cause real-world compensation. Experiment 3 further showed significant compensation. This experiment showed that the time between comparisons when comparing stimuli in a real-world situation is partly behind real-world compensation. Further, the instructions given to listeners and the task-up set were not behind real-world compensation. Compensation appears to be a genuine listening phenomenon.

Therefore, the work in this chapter provided answers to Question 1a. Compensation for loudspeakers and rooms was shown to occur in real-world listening in 3 separate tests. It
was moderate to large and highly significant. These results, together with Olive et al’s study, show compensation for loudspeakers and rooms occurs when listening to music and speech. It was shown throughout these studies that the effect of this compensation is to reduce the perception of the channel when the listener is rating sounds for timbral preference. This has the effect of making the same sounds produced through different channels more timbrally similar (i.e. timbral constancy is achieved). The sound produced through different channels is heard to be more like its canonical form and in a relatively channel-free manner. Compensation was shown by translational shifts of ratings toward the perceptual midpoint with real-world listening. Compensation for loudspeakers and rooms was been adequately confirmed by the work presented. Its extent was been measured and its nature shown.

The work in this chapter also contributed to answering Question 1b. In a preliminary investigation it was shown that compensation was not due to longer listening between approximately 25 s to 4 minutes, but may instead be due to continuous listening prior to 25 s. The effect of longer listening was not tested further because this was not possible in that study. The effect of time-gap on compensation was tested. Time gaps between stimuli present in real-world listening where shown to explain compensation for the room factor but results were inconclusive for the loudspeaker factor. It was concluded that time gaps between stimuli partially explain real-world compensation. The effect of instructions were tested. It was shown that bias due to instructions did not cause compensation. Compensation is not an artefact of the specific task. Further, compensation does not appear to be due to attention. It was also apparent that non-time-gap sensitive mechanisms cause the compensation that remains after the time gaps and bias due to instructions are removed. Therefore, it may be concluded that the mechanisms of real-world compensation were determined to some extent by the work presented in Chapter 3.

Chapter 4 did not aim to provide answers to the research questions but set out a research process to do this. It was proposed that a literature review of compensation mechanisms should be undertaken to further determine the cause of real-world compensation. To limit the extent of this work it was decided that only the time-gap sensitive component of compensation shown in Experiment 3 would be investigated, but that mechanisms behind this component of compensation should be determined.
for both speech and non-speech listening.

Chapter 5 presented a literature review to determine potential mechanisms of the time-gap sensitive component of real-world compensation. A number of mechanisms were put forward that have been shown in psychoacoustic laboratory tests to cause compensation for the spectral colouration caused by transmission channels (and the vocal tract and phonetic context in speech perception). It was noted that they may all be potential mechanisms of the time-gap sensitive component of real-world compensation, providing: they are time-gap sensitive; criteria can be met to show that they have the nature and time course to explain real-world compensation and; are unique mechanisms, not manifestations of another (i.e. if Questions A-C of the research process set out in Chapter 4 can be answered in the affirmative). The literature review provided a large amount of information on a number of possible mechanisms, allowing for their nature and time course to be assessed. Of all the mechanisms discussed only two were shown to be potential mechanisms of the time-gap sensitive component of real-world compensation for loudspeakers and rooms: the enhancement effect and the spectral compensation effect. These mechanisms have not been shown to occur with non-speech listening. Therefore, it was not possible to conclude that these mechanisms are potential mechanisms of the time-gap sensitive component of compensation when listening to non-speech. It was concluded that it was necessary to confirm these mechanisms with non-speech using laboratory tests.

The research in this chapter did not aim to answer Research Question 1a as real-world compensation and its extent were not specifically examined in the literature review. Research Question 1b was further answered. Evidence of two specific potential mechanisms of the time-gap sensitive component of real-world compensation was provided. Research Question 1b is therefore adequately answered for speech listening, but it should also be confirmed that these mechanisms actually cause real-world compensation (i.e. question 1c should be addressed). Further there may be other, currently unknown, compensation mechanisms that also contribute to the time-gap sensitive competent of compensation in real-world listening and real-world tests would show this. These tests are not conducted as part of this thesis. It was concluded that it should also be shown that these mechanisms occur with non-speech listening. This would allow for a determination of whether general auditory mechanisms explain
the time-gap sensitive component of compensation. It was concluded that this would be examined in this thesis. Although research Question 1b was answered for the non-time-gap sensitive mechanisms (for speech) the mechanisms behind the non-time-gap sensitive portion of compensation have not been properly considered. Further work could aim to determine the mechanisms behind this component of compensation.

Chapter 6 examined whether the enhancement effect and the spectral compensation effect occur with non-speech. It also set up test material for any further testing of these mechanisms that might be required. Experiment 4 confirmed that shifts indicative of compensation via one of these mechanisms occur in tests with music and noise. Experiment 5 used crossaural presentation to determine whether central or monaural effects caused the shifts. For noise only a monaural effect caused shifts, therefore only the enhancement effect occurs with noise. For music both a central component and a monaural component were seen, suggesting both the spectral compensation effect and the enhancement effect contribute to compensation with the music seen in Olive et al’s original test. Therefore, it can be concluded that these mechanisms are potential mechanisms of the time-gap sensitive component of compensation seen in real-world listening to speech and in Olive et al’s original test with music.

The chapter did not aim to answer Question 1a. Question 1b was further answered by showing the enhancement and spectral compensation effects are potential mechanisms of compensation (specifically the time-gap sensitive component) when listening to non-speech. Both may cause real-world compensation when listening to music but only enhancement occurs when listening to noise.

It is possible to conclude that Research Questions 1a and 1b have been answered through the work in this thesis. Real-world compensation is shown to reduce the perceived effect of loudspeakers and rooms, making the same sound heard through different channels more timbrally similar, resulting timbral constancy. The extent of this compensation is moderate to large. The potential mechanisms behind the time-gap sensitive component of this real-world compensation are the spectral compensation effect and the enhancement effect, when listening to speech and non-speech. Further work is necessary to confirm the role of these mechanisms in real-world compensation.
and to establish the mechanisms behind the non-time-gap sensitive component of compensation.

7.3.1 Original contributions to knowledge

- Confirmed a source of compensation for transmission channels (loudspeakers and rooms) occurs with real-world listening but not laboratory listening.

- Established the hypothesis that this real-world compensation is due to longer continuous listening and attention (time gap already hypothesised to be cause) and established that time gaps cause real-world compensation.

- Established that real-world compensation does not occur due to instructions or the task set up causing bias.

- Conducted the first comprehensive literature review of research in this field.

- Established to a high degree of certainty which compensation mechanisms that have previously been researched are *unique*.

- Put forward a hypothesis that two compensation mechanisms, enhancement and spectral compensation, have the nature and time course to cause to real-world compensation (specifically the time gap sensitive component of this).

- Established shifts indicative of enhancement and spectral compensation in a test using only *natural* non-speech musical stimuli, so effect is not *speech* specific.

- Established that only the enhancement effect occurs in with noise stimuli. The spectral compensation effect does not occur with noise.

- Established that both enhancement and spectral compensation occur with the music tested. Therefore, both have the potential to contribute to the time-gap sensitive component of real-world compensation seen in Olive et al’s experiment.

- Fully discussed the involvement of memory in channel compensation for the first time.
Appendix A

Participant instructions

The following written instructions were the standard instructions issued to participants in Experiments 1-3. A further description of instructions is given in Sections 3.2.5, 3.9 and 3.16 for Experiments 1-3 respectively.

You will hear 27 versions of an 8 bar female vocal track. You are asked to rate each one for preference. You will hear 3 versions (A, B, C) presented at one time. You can switch between the 3 versions by pressing the A, B, C buttons above the slider.

You must make a rating for your preference for each version using the slider to mark a preference score ranging from 1 - 100. A score of 10 should be used to represent ‘really dislike’ and 90 ‘really like’. A score of 50 should be used to represent ‘neither like nor dislike’. To mark a strong difference between stimuli, use an interval of 20 or more points. To mark a moderate difference in preference use at least 10 points difference and to mark a slight preference use at least 5 points.

The music will loop continuously but you may press pause at any time.

You may also stop the track at any time and start it again, to listen from the beginning.

You can only move the slider to make a rating for a track while it is currently playing.
Appendix B

Binaural recordings

Figure B.1 displays the apparatus set-up for the measurement of binaural room impulse responses used to make stimuli for Experiments 1-3, described in Chapter 3. A full description of the binaural recording process is given in Section 3.2.3.

A Cortex Instruments, Manikin Mk2 was used as a head and torso simulator (HATS). Sine sweeps were generated using Adobe Audition v2.0, played through each loudspeaker in each room, and recorded using microphones placed within each ear of the HATS.

Figure B.1: Recording apparatus.
Appendix C

Task Interface

Figure C.1 shows the graphical user interface for the page based experiments described in Chapter 3 (Experiments 1-3). A full description of the interface is given in Section 3.2.4.

Figure C.1: Task interface.
Appendix D

Photographs

Figures D.1, D.1 and D.3 show photographs of the rooms used to make binaural recordings of loudspeakers playing in rooms, and the binaural recording set-up. See Section 3.2.3 for a full description of the room characteristics 3.2.3.
Figure D.3: Room3.
Appendix E

Reliability Analysis

Figures E.1 to E.8 show reliability analyses for Experiment 1 (see Section 3.4.1). The results are shown for each participant for each loudspeaker/room combination and each repeat: BO=LS1, AE=LS2, BW=LS3; OFF=Office, Cont2=Studio Control Room, Tb7=ITU-R BS.1116 Listening Room; 1=Repeat1, 2=Repeat2, 3=Repeat3.

Graphs for each participant show preference score by repetition for each LS/RM combination. Cronbach’s alpha measures reliability of participant across the experiment. Correlations of above .7 = good reliability. Data was analysed for the loudspeaker condition only.
Figure E.1: Reliability analysis of Participant 1. Cronbach’s alpha = .789

Figure E.2: Reliability analysis of Participant 2. Cronbach’s alpha = .264
Figure E.3: Reliability analysis of Participant 3. Cronbach’s alpha = .246

Figure E.4: Reliability analysis of Participant 4. Cronbach’s alpha = .823
Figure E.5: Reliability analysis of Participant 5. Cronbach’s alpha = .001

Figure E.6: Reliability analysis of Participant 6. Cronbach’s alpha = .771
Figure E.7: Reliability analysis of Participant 7. Cronbach’s alpha = .864

Figure E.8: Reliability analysis of Participant 8. Cronbach’s alpha = .835
Appendix F

Publications

The following published research papers by this author contain work relating to this thesis:

Pike, C., Brookes, T., and Mason, R. (2013). Auditory adaptation to loudspeaker and listening room acoustics, 135th Convention of the Audio Engineering Society, New York, 17-21 Oct, preprint 8917. This paper presents results from Experiments 2 and 3 in this thesis. This work is also presented in Chapter 3 of this thesis.

Pike, C., Mason, R., and T. Brookes. (2014). The effect of auditory memory on the perception of timbre, 136th Convention of the Audio Engineering Society, Berlin, Germany. 26-29 April, preprint 9028. This paper presents results of a side experiment investigating the effect of timbral memory decay, which was not conducted as part of this thesis but is referenced in Chapter 6.

Pike, C., Brookes, T., and Mason, R. (2014). Auditory compensation for spectral colouration, 137th Convention of the Audio Engineering Society, Los Angeles, USA. 9-12 Oct, preprint 9138. This paper presents the results from Experiment 4 in this thesis, which is a psychoacoustic laboratory experiment showing extrinsic compensation with non-speech stimuli. This work is presented in Chapter 6 of this thesis.
Appendix G

Glossary

The following terms and abbreviations are used in this thesis:

Adaptation (perception): this term is generally used to describe the effect of the general perception of a stimulus changing over time or depending on the environment. Specifically, this process describes a decrease in sensitivity to some aspect of the perception—the aspect that is becoming adapted—and this results in other aspects of the perception becoming relatively more prominent, hence the overall change in perception (see also simple neural adaptation).

ANOVA: Analysis of Variance. A statistical analysis technique for quantifying the the effect of a variable/s or ‘factor/s’ within a dataset relative to amount of variation due to factors not included in the analysis or random errors within that data set.

Canonical form: The idealised or standard form of an object.

Categorical perception (CP): the effect of perceiving stimulus attributes which vary continuously as being a members of discrete categories. By this process sensitivity to within-category variation is lost or reduced but sensitivity to between-category variation remains. This process can result in target cues to stimulus identification being perceived categorically and the categorical perception of objects that are identified by those cues.

Categorisation: the process of assigning stimuli to categories using a number of
attributes that signal category membership. When sufficient cues signal membership of a category, categorisation takes place. Once categorisation has occurred listeners may become less sensitive to within-category variation.

**CGS:** Category guided separation. A process by which listeners are able to separate two stimuli with different spectra, such as two intermixed vowel sounds, by using templates which are stored in memory to identify and perceptually separate each.

**Central processing:** this term is used to describe processes occurring beyond the cochlear nuclei, where stimuli from both ears meet. Central processes occur at a higher level in the stimulus-to-perception chain. The term central often refers to cognitive processes but may be used to refer to any physiological processes occurring at or later than the cochlear nuclei. Central processes can be crossaural.

**CI:** Confidence interval. A measure of the bounds within which a parameter lies with 95%, or some other level of, certainty.

**Cognitive processes:** this term describes complex executive and intellectual functions such as memory, learning, and attention which occur at a high level in the stimulus-to-perception chain. The exact physiological underpinnings of these processes are not yet known.

**CP:** Categorical perception (see categorical perception).

**Crossaural effect:** this term describes an effect where a sound presented at one ear affects perception of a sound presented to another ear.

**DSV:** Double Spectrum Vowel. This was a test stimulus used in a study by Summerfield and Assmann (1989), which consisted of a harmonic complex with several components raised in amplitude to form formant frequencies of a single vowel, plus components raised in amplitude to form formant frequencies of another vowel, thus producing a ‘double vowel’. The second vowel represented the effect of channel colouration.

**Enhancement:** this term describes a situation in which the perception of the difference in spectrum between two sounds is augmented, or the specific effect seen
in studies by Summerfield et al. and others which results in this perception. This effect may be caused by the same mechanisms responsible for timbral constancy.

**FSV:** Flat spectrum vowel. This describes the perception of a test stimulus used in studies by Summerfield et al. (Summerfield and Assmann 1989, Summerfield et al. 1984, 1987). The flat spectrum test stimulus consisted of a flat harmonic complex (or complex with a gradual roll-off). After the presentation of a precursor stimulus (the same harmonic complex with the amplitude of components representing a vowel sound reduced), the perception of a vowel in the flat spectrum test stimulus was induced, creating a ‘flat spectrum vowel’.

**HATS:** Head And Torso Simulator. This term describes a manikin with built-in ear simulators that provides a realistic reproduction of the acoustic properties of an average adult human head and torso.

**LTAS:** the Long Term Average Spectrum. This describes the magnitude spectrum averaged over a period of time.

**LS:** Loudspeaker.

**Monaural effect:** this term describes an effect where a sound presented to one ear does not affect the perception of a sound presented to another ear.

**MOC:** Medial Olivocochlear. The MOC system involves an efferent feedback system, via the Olivocholear nerve bundle (OCB), thought to be involved in cochlear protection against loud sounds and the detection and discrimination of sounds in noise. The OCB begins in the superior olivary complex and projects to the ipsilateral and contralateral cochleae. This process innervates the outer hair cells in the cochlea and through a series of adjustments improves the neural representation of sound by controlling the encoding of the signal’s dynamic range. This system may also be involved in hearing in the presence of reverberation and spectral colouration.

**Perceptual constancy (Auditory):** the effect of the perceived attributes of a sound remaining stable despite modifications to the objective characteristics of the sound caused by variations in environmental factors (e.g. room reflections, background noise). This term may also be used to describe the process by which this effect occurs (see
also timbral constancy).

**Perceptual constancy (Colour):** the effect of the colour of a object remaining perceptually stable despite modifications to the spectrum of light reflected from that object due to variations in the spectrum of illumination. This term may also be used to describe the process by which this effect occurs.

**Perceptual Magnet Effect:** this describes a less extreme form of categorical perception whereby listeners show a tendency to perceive objects categorically through increased sensitivity to stimuli differences at the edge of category boundary and decreased sensitivity to stimuli differences near the nucleus of a category, but not complete loss of sensitivity to within-category variation.

**Peripheral:** in speech psychophysics the term peripheral is used to describe processes occurring below the cochlear nuclei and therefore at a lower level in the stimulus-to-perception chain. Peripheral processes are only monaural.

**Physiological processes:** these processes are processes for which the biological (mechanical, electrical, chemical, or ‘neural’) underpinnings are well described. They occur at all levels of the stimulus-to-perception chain. At lower levels, physiological processes are particularly well understood. See also ‘Cognitive’.

**PME:** the Perceptual Magnet Effect (see Perceptual Magnet Effect).

**Psychometric function:** this is term is usually used to describe a graphical function which describes the response of a subject to a physical stimulus that changes along a one-dimensional continuum. The probability of a particular binary response choice is displayed on the y axis and the stimulus value is displayed on the x axis. These functions are usually obtained over multiple repeated listens to sounds from the continuum and are frequently sigmoid shaped where a change in perception occurs gradually with stimulus change.

**PSV:** Peaked Spectrum Vowel. This was a test stimulus used in studies by Summerfield et al. (Summerfield and Assmann 1989, Summerfield et al. 1987), which consisted of a harmonic complex with several components raised in amplitude to form formant frequencies of a single vowel, thus creating a synthetic vowel stimulus.
**Q-Q plot:** also known as a Quantile-Quantile plot. This plot describes the variation of individual data points from given distribution such as that of a second variable, or a hypothetical distribution such as the normal distribution. The proportion of data points at given values is assessed against the proportion expected according to the second variable or hypothetical distribution.

**Real-world listening:** this term describes a situation that is more similar to listening in everyday life than in laboratory experiments. The main feature of this is that listening is over a longer time course than in the laboratory but other differences include listening to multiple stimuli at the same time.

**RM:** Room.

**RT60:** a measure of room reverberation decay. RT60 shows the amount of time necessary for the level of room reverberation to reduce by 60 decibels.

**Spectro-temporal variation:** this term describes amplitude variation in spectral components of a sound over time. It may be used to describe only such variations that are perceptible within the resolution of the hearing system.

**Stimulus-to-perception chain:** this describes the chain of events involved in listening. This chain begins with the physical sound being modified by the outer ear and results in cognitive processing whereby the sound is recognised and then semantic processing whereby the sound is given meaning and emotional significance.

**SSA:** Stimulus specific adaptation. This is adaptation that occurs in populations of neurons at multiple levels in the hearing system. Responses in neurons with continuous or repeated stimulation is reduced but returns to baseline with a new or ‘deviant’ stimulus.

**Simple neural adaptation:** This describes the decrease in neural firing rate in response to continuous stimulation by a physical stimulus. This can result in the perception of that stimulus becoming ‘fatigued’ or reduced in strength. Where this occurs for some aspects of a stimulus but not others this changes the general perception of the stimulus, as the non-adapted components have increased perceptual weight. (see also adaptation (perception))
**Timbral Constancy:** this describes the effect of the timbre of a sound source remaining constant despite modifications to the objective characteristics of the sound caused by environmental factors. In particular in Chapter 2 onwards of this thesis this term is restricted to describing the effect of timbre remaining constant despite modification to the spectral envelope of a sound caused by spectral colouration by the transmission channel (and in speech perception caused by the phonetic context).

**Time-gap sensitive compensation:** this describes a compensation process that occurs to a greater or lesser degree depending on the time between the stimuli affected.

**VT:** the vocal tract (see Vocal Tract).

**Vocal tract:** the sound producing cavity which lies above the larynx. It includes the laryngeal cavity, the pharynx, the oral cavity and the nasal cavity. The shape of this is modified in speech production to produce various speech sounds.
References


hall’ Text of an oral presentation at Parc de la Villette, Paris; part of a series of lectures entitled ‘Acoustique, Musique, Espaces’.


Carlson, R., Fant, G. & Granström, B. (1975), *Two formant models, pitch and vowel*
perception., Academic Press, chapter In Auditory analysis and perception of speech, pp. 55–82.


Fant, G. (1973), *Speech sounds and features*, MIT, Cambridge M.A.


REFERENCES


REFERENCES


Lloyd, R. (1890a), Some research into the nature of vowel-sound, Turner and Dunnett, Liverpool, England.

Lloyd, R. (1890b), ‘Speech sounds their nature and causation i’, Phonetische Studien 3, 251–278.


REFERENCES


Ohl, F. & Scheich, H. (1997), ‘Orderly cortical representations of vowels based on


variability, speaker rate and perceptual learning’, *Speech Communication* pp. 109–125.


processes in speech perception: evidence from a dichotic study’, *Journal of Cognitive Psychology* 2, 455–466.


Toole, F. (2008), *Sound reproduction: the acoustics and psychoacoustics of loudspeakers and rooms*, Focal Press.


redundancy: intelligibility of sentences heard through narrow spectral slits’, *Perception and Psychophysics* 57, 175–182.


