Developing a Graphical Language to Represent Listeners' Experiences of Spatial Attributes in Reproduced Sound

by

Natanya Ford

Thesis submitted in fulfilment of the requirements of the degree of Doctor of Philosophy

University of Surrey
Department of Music & Sound Recording
School of Performing Arts

February 2005

© Natanya Ford 2005
Spatial attributes of reproduced sound have typically been described using primarily verbal descriptors. To gain an alternative perspective, the development of a communication medium is discussed which enables listeners’ auditory spatial experiences to be represented graphically.

Graphical descriptions of auditory spatial experiences were sought from listeners when multichannel audio was reproduced within an automotive setting. In order to avoid biasing results by adopting traditional assumptions relating to spatial audio evaluation, the listener was acknowledged as the originator of meaning and valid results sought by the exploration and understanding of the listeners’ individual experiences.

Because experiences are pre-linguistic and distinct from language, a communication medium was required for the structuring and representation of listeners’ auditory spatial experiences in order that these could be understood by the researcher. A descriptive graphical language (GAL) was systematically developed and evaluated for this purpose.

A series of investigations was conducted during which GAL evolved from a system of individually elicited descriptors to a universal language of graphical terms, capable of representing the salient spatial experiences of listeners with varying levels of critical listening expertise.

At each successive stage in GAL’s development, ambiguities which could impede the researcher’s understanding of listeners’ graphically represented spatial experiences were identified and minimised. A graphical model was presented for visualising instances where ambiguities occurred in the descriptive process. One notable anomaly was found to be related to the provision of unsuitable written instruction in listening investigations. The effect of this ambiguity was minimised by listeners developing their own verbal descriptors to accompany their graphical language.

When a mutual language - developed by the listeners themselves - was evaluated, an unambiguous route through the descriptive process from listener to researcher was identified. Within the context of the research, it was concluded that this particular language enabled the researcher to acquire a valid understanding of listeners’ auditory spatial experiences. It was further concluded that the method employed in the development of the language (an iterative process comprising descriptor elicitation, clarification, development, evaluation and ultimately validation) could be used in alternative contexts within subjective audio evaluation.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter 1 Introduction</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Introduction</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Historical perspectives – setting the research context</td>
<td>4</td>
</tr>
<tr>
<td>1.2.1 Perspectives on the use of listeners in the evaluation of reproduced audio</td>
<td>4</td>
</tr>
<tr>
<td>1.2.2 Methods for obtaining meaningful results from subjective evaluation</td>
<td>6</td>
</tr>
<tr>
<td>The influence of the loudspeaker and listening location in the listening room</td>
<td>6</td>
</tr>
<tr>
<td>The choice of stimuli for subjective audio evaluation</td>
<td>6</td>
</tr>
<tr>
<td>The influence of loudness</td>
<td>7</td>
</tr>
<tr>
<td>Bias introduced by the experimenter and experiment process</td>
<td>7</td>
</tr>
<tr>
<td>The importance of blind testing in subjective evaluation</td>
<td>8</td>
</tr>
<tr>
<td>1.2.3 Conducting dependable evaluation using human subjects</td>
<td>9</td>
</tr>
<tr>
<td>Devising an appropriate scale for obtaining accurate responses</td>
<td>9</td>
</tr>
<tr>
<td>Defining experience</td>
<td>10</td>
</tr>
<tr>
<td>Selecting listeners for subjective listening evaluation</td>
<td>11</td>
</tr>
<tr>
<td>Benefits of experience and expertise</td>
<td>11</td>
</tr>
<tr>
<td>Training listeners</td>
<td>12</td>
</tr>
<tr>
<td>Concluding remarks</td>
<td>13</td>
</tr>
<tr>
<td>1.2.4 An alternative perspective on the measurement of meaning</td>
<td>15</td>
</tr>
<tr>
<td>The problem of neglecting an individual’s perception</td>
<td>15</td>
</tr>
<tr>
<td>A counter-argument for listener selection and training</td>
<td>16</td>
</tr>
<tr>
<td>Positioning meaning with the subject rather than the object of subjective evaluation</td>
<td>18</td>
</tr>
<tr>
<td>Practical considerations for a more qualitative understanding of listeners’ experiences</td>
<td>20</td>
</tr>
<tr>
<td>1.2.5 Communicating experiences</td>
<td>23</td>
</tr>
<tr>
<td>The definition of a linguistic sign</td>
<td>24</td>
</tr>
<tr>
<td>The arbitrary nature of signs</td>
<td>24</td>
</tr>
<tr>
<td>The context dependency of language in communication</td>
<td>25</td>
</tr>
<tr>
<td>The idiosyncrasies and inadequacies of language</td>
<td>25</td>
</tr>
<tr>
<td>Personal meaning in language</td>
<td>26</td>
</tr>
<tr>
<td>1.2.6 Eliciting verbal descriptions from participants in subjective evaluation</td>
<td>29</td>
</tr>
<tr>
<td>The definition of a spatial attribute</td>
<td>30</td>
</tr>
<tr>
<td>Eliciting verbal spatial attribute descriptions from listeners</td>
<td>31</td>
</tr>
<tr>
<td>RGT and the use of elicited terminology</td>
<td>32</td>
</tr>
<tr>
<td>The use of Descriptive Analysis in spatial audio evaluation</td>
<td>33</td>
</tr>
<tr>
<td>Concluding remarks</td>
<td>35</td>
</tr>
<tr>
<td>1.2.7 Alternative methods for representing spatial percepts</td>
<td>37</td>
</tr>
<tr>
<td>The egocentric nature of directional spatial experiences</td>
<td>37</td>
</tr>
<tr>
<td>Eliciting directional information from participants using non-verbal methods</td>
<td>37</td>
</tr>
<tr>
<td>Sources of ambiguity in non-verbal response methods</td>
<td>38</td>
</tr>
<tr>
<td>Translating egocentric spatial experiences into graphical representations</td>
<td>40</td>
</tr>
<tr>
<td>Using a graphical medium in subjective audio evaluation</td>
<td>42</td>
</tr>
<tr>
<td>1.3 Summary</td>
<td>44</td>
</tr>
<tr>
<td>1.4 Statement of research objectives</td>
<td>47</td>
</tr>
<tr>
<td>1.5 Evaluating the effectiveness of the descriptive graphical language</td>
<td>49</td>
</tr>
<tr>
<td>1.6 Original contributions of the author</td>
<td>51</td>
</tr>
<tr>
<td>1.7 Thesis overview</td>
<td>52</td>
</tr>
</tbody>
</table>

| Chapter 2 A summary of three initial investigations | 55 |
| 2.0 Chapter overview | 56 |
| 2.1 Introduction | 57 |
| 2.2 A pilot study into the graphical description of spatial attributes | 59 |
| 2.2.1 Details of the investigation setting | 59 |
| Stimuli | 59 |
| Loudspeaker and listening locations | 60 |
## Table of contents

1. The graphical representation of listeners' auditory spatial experiences

2.2.2 Analysis of responses

2.2.3 Summary of findings

2.3 A formal study of graphical description

2.3.1 Details of the investigation setting

2.3.2 Analysis of graphical descriptions

2.3.3 Summary of findings

2.4 The further development of a descriptive graphical language (GAL)

2.4.1 Details of the investigation setting

2.4.2 Analysis of listeners' responses

2.4.3 Summary of findings

2.5 Conclusions

2.5.1 Summary of three initial investigations

2.5.2 The necessity for further work

2.6 Chapter summary

Chapter 3 Clarifying individual listeners' descriptions & developing U-GAL

3.0 Chapter overview

3.1 Introduction

3.1.1 Identifying sources of ambiguity in graphical responses

3.2 Clarification process

3.2.1 Elicitation of individual graphical descriptors

3.3 Findings of the clarification study

3.3.1 Clarifying the attribute individual instrument width

3.3.2 Focus and localisation

3.3.3 Width and envelopment – a continuum?

3.3.4 The description of alternative attributes

3.3.5 Attributes specific to ensembles

3.4 Summary of the findings from the clarification process

3.5 Rationale for the continued development of GAL
<table>
<thead>
<tr>
<th>Chapter 5 Conclusions</th>
<th>190</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.0 Chapter overview</td>
<td>191</td>
</tr>
<tr>
<td>5.1 Conclusions</td>
<td>192</td>
</tr>
<tr>
<td>5.1.1 Influences of the research context and other considerations</td>
<td>193</td>
</tr>
<tr>
<td>Considerations regarding the research ethos</td>
<td>194</td>
</tr>
<tr>
<td>Considerations regarding the use of language</td>
<td>194</td>
</tr>
<tr>
<td>The balance between retaining meaning &amp; obtaining understanding</td>
<td>194</td>
</tr>
<tr>
<td>5.1.2 Concluding remarks</td>
<td>195</td>
</tr>
<tr>
<td>5.2 Further development of the descriptive language</td>
<td>196</td>
</tr>
</tbody>
</table>

Appendices | 197
References | 283
LIST OF FIGURES

Figure 1.2.1 Shaw and Gaines, model of conceptual structures (1995). Cited Berg (2002)
Figure 1.4.1 Graphical language (GAL) development model
Figure 2.2.1 Loudspeaker and listening locations
Figure 2.2.2 Guide sheet for graphical responses
Figure 2.2.3 Illustration of potential drawing style
Figure 2.2.4 - 2.2.6 Listeners individual depictions of the folk rock stimulus from seat 2
Figure 2.2.7 - 2.2.9 All listeners, all seats, responses for principal instruments within different ensembles
Figure 2.2.10 Listener 2 depiction of the choral music from seat 2
Figure 2.2.11 Listener 3 depiction of choral music from seat 2
Figure 2.2.12 All listeners, all seats, depictions of the voice from the choral ensemble
Figures 2.2.13 - 2.2.15 Total image plots for different ensembles, all listeners, all seats.
Figure 2.2.16 Measuring ensemble location
Figure 2.2.17 Establishing measurable characteristics of ensembles
Figure 2.2.18 All listeners descriptions of ensemble width (measured in mm)
Figure 2.2.19 All listeners descriptions of ensemble depth (measured in mm)
Figure 2.2.20 All listeners descriptions of ensemble skew (measured in degrees)
Figure 2.3.1 Architecture of the audio reproduction system within the vehicle
Figure 2.3.2 Default audio system setting
Figures 2.3.4 - 2.3.7 Individual responses from listeners to programme material reproduced using the stereo setting
Figure 2.3.8 Exemplary driver seat description
Figure 2.3.9 Exemplary centre seat description
Figure 2.3.10 Exemplary rear left seat description
Figure 2.3.11 Listener 3 description of extract 1 (1st time)
Figure 2.3.12 Listener 3 description of extract 1 (2nd time)
Figure 2.3.13 Descriptions of percussion from all listening locations over stereo and processed settings
Figure 2.3.14 Descriptions of the cello from all listening locations over stereo and processed settings
Figure 2.3.15 Means and 95% confidence intervals for ensemble width depicted from different listening locations over different reproduction system settings
Figure 2.3.16 Means and 95% confidence intervals for ensemble location depicted from different listening locations over different reproduction system settings
Figures 2.3.17 - 2.3.22 Descriptive differences when system setting changed from stereo (s) to processed (p) mode
Figure 2.4.1 Graphical response sheet
Figure 2.4.2 Listener 5 percussion (optimal system setting)
Figure 2.4.3 Listener 3 percussion (optimal system setting)
Figure 2.4.4 Listener 2 cello (reduced bandwidth centre loudspeaker)
Figure 2.4.5 Listener 9 early depiction of percussion (optimal system setting)
Figure 2.4.6 Listener 9 later depiction of cello (optimal system setting)
Figure 2.4.7 Listener 7 cello (phantom centre)
Figure 2.4.8 Listener 17 percussion (optimal system setting)
Figure 2.4.9 Listener 11 cello (optimal system setting, depiction from the passenger’s seat)
Figure 2.4.10 Listener 11 percussion (phantom centre, depiction from the passenger’s seat)
Figure 2.4.11 Listener 8 cello (optimal system setting)

1 Appendices have their own table of contents listing any figures and tables included within their pages
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4.12</td>
<td>Listener 13 cello (optimal)</td>
<td>91</td>
</tr>
<tr>
<td>2.4.13</td>
<td>Listener 13 cello (optimal repeat)</td>
<td>91</td>
</tr>
<tr>
<td>2.4.14</td>
<td>Means and 95% confidence intervals for instrument width depicted</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td>for different programme items over different reproduction system settings</td>
<td></td>
</tr>
<tr>
<td>2.4.15</td>
<td>Overlaid responses for the cello (optimal system setting)</td>
<td>94</td>
</tr>
<tr>
<td>2.4.16</td>
<td>Overlaid responses for the percussion (optimal system setting)</td>
<td>94</td>
</tr>
<tr>
<td>2.4.17</td>
<td>Overlaid responses for the cello (phantom centre)</td>
<td>94</td>
</tr>
<tr>
<td>2.4.18</td>
<td>Overlaid responses for the percussion (phantom centre)</td>
<td>94</td>
</tr>
<tr>
<td>2.4.19</td>
<td>Means and 95% confidence intervals for instrument location depicted</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>from different listening locations for different reproduction system settings</td>
<td></td>
</tr>
<tr>
<td>2.5.1</td>
<td>GAL development model stage I: The initial GAL investigations</td>
<td>102</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Graphical response sheet</td>
<td>114</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Listener 2 initial description (CPVR)</td>
<td>116</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Listener 2 repeat description (CPVR)</td>
<td>116</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Listener 16 initial description (Perc C)</td>
<td>116</td>
</tr>
<tr>
<td>3.3.4-7</td>
<td>Selected descriptions of the central percussion stimulus</td>
<td>117</td>
</tr>
<tr>
<td>3.3.8-13</td>
<td>Selected descriptions of the central cello stimulus, widened by</td>
<td>117</td>
</tr>
<tr>
<td></td>
<td>artificial stereo reverberation</td>
<td></td>
</tr>
<tr>
<td>3.3.14</td>
<td>'depth of field'</td>
<td>120</td>
</tr>
<tr>
<td>3.3.15</td>
<td>'more than a thin line of sound'</td>
<td>120</td>
</tr>
<tr>
<td>3.3.16</td>
<td>'further away'</td>
<td>120</td>
</tr>
<tr>
<td>3.3.17</td>
<td>'taking a back seat to other instruments'</td>
<td>120</td>
</tr>
<tr>
<td>3.3.18</td>
<td>'spread / projection'</td>
<td>121</td>
</tr>
<tr>
<td>3.3.19</td>
<td>'spread across'</td>
<td>121</td>
</tr>
<tr>
<td>3.3.20</td>
<td>'sound coming across'</td>
<td>121</td>
</tr>
<tr>
<td>3.3.21</td>
<td>'wrapped around'</td>
<td>121</td>
</tr>
<tr>
<td>3.3.22</td>
<td>Listener 17 'listening from a distance'</td>
<td>122</td>
</tr>
<tr>
<td>3.3.23</td>
<td>Listener 20 'width is all over the front'</td>
<td>122</td>
</tr>
<tr>
<td>3.3.24</td>
<td>Listener 15 'coming from all around me'</td>
<td>122</td>
</tr>
<tr>
<td>3.3.25</td>
<td>Listener 15 'not coming up to me'</td>
<td>122</td>
</tr>
<tr>
<td>3.3.26</td>
<td>'well defined instruments in space'</td>
<td>124</td>
</tr>
<tr>
<td>3.3.27</td>
<td>'clear separation between instruments'</td>
<td>124</td>
</tr>
<tr>
<td>3.3.28</td>
<td>'joined'</td>
<td>124</td>
</tr>
<tr>
<td>3.3.29</td>
<td>'sounds can be distinguished'</td>
<td>124</td>
</tr>
<tr>
<td>3.3.30</td>
<td>One or more ensembles?</td>
<td>124</td>
</tr>
<tr>
<td>3.3.31</td>
<td>Two or three ensembles?</td>
<td>124</td>
</tr>
<tr>
<td>3.3.32</td>
<td>Two or three ensembles?</td>
<td>124</td>
</tr>
<tr>
<td>3.3.33</td>
<td>Single ensemble?</td>
<td>124</td>
</tr>
<tr>
<td>3.4.1</td>
<td>GAL development model stage II: Known ambiguities following the</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td>clarification of listener descriptions</td>
<td></td>
</tr>
<tr>
<td>3.6.1</td>
<td>Graphical representations of area &amp; focal point as developed by the</td>
<td>138</td>
</tr>
<tr>
<td></td>
<td>five individual panels</td>
<td></td>
</tr>
<tr>
<td>3.7.1</td>
<td>Graphical language for describing a single instrument</td>
<td>139</td>
</tr>
<tr>
<td>3.7.2-7</td>
<td>Representing audio images</td>
<td>141</td>
</tr>
<tr>
<td>3.8.1</td>
<td>GAL development model stage III: Potential ambiguities following the</td>
<td>144</td>
</tr>
<tr>
<td></td>
<td>development of U-GAL</td>
<td></td>
</tr>
<tr>
<td>4.2.1</td>
<td>Grid at start of investigation</td>
<td>154</td>
</tr>
<tr>
<td>4.2.2</td>
<td>U-GAL (blank) response page</td>
<td>154</td>
</tr>
<tr>
<td>4.2.3</td>
<td>Grid displaying a number of completed runs</td>
<td>154</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Exemplary use of U-GAL descriptors: Focused cello with a separate</td>
<td>157</td>
</tr>
<tr>
<td></td>
<td>feeling of space</td>
<td></td>
</tr>
<tr>
<td>4.3.2</td>
<td>Exemplary use of U-GAL descriptors: Unfocused cello with a</td>
<td>157</td>
</tr>
<tr>
<td></td>
<td>surrounding feeling of space</td>
<td></td>
</tr>
<tr>
<td>4.3.3</td>
<td>Listener 1 depiction of (Cello C)</td>
<td>158</td>
</tr>
<tr>
<td>4.3.4</td>
<td>Listener 1 depiction of ensemble (CPVR)</td>
<td>158</td>
</tr>
<tr>
<td>4.3.5</td>
<td>Listener 2 depiction of (Cello C)</td>
<td>158</td>
</tr>
<tr>
<td>4.3.6</td>
<td>Listener 2 depiction of ensemble (CPVR)</td>
<td>158</td>
</tr>
<tr>
<td>4.3.7</td>
<td>Listener 3 depiction of ensemble (VCPW) letters outside of FOCAL</td>
<td>158</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>4.3.8</td>
<td>Listener 6 depiction of ensemble (CPVN) letters outside of FOCAL</td>
<td></td>
</tr>
<tr>
<td>4.3.9</td>
<td>Listener 13 depiction of ensemble (VCPL) consistency in letter use</td>
<td></td>
</tr>
<tr>
<td>4.3.10</td>
<td>Listener 18 depiction of ensemble (VCPL) consistency in letter use</td>
<td></td>
</tr>
<tr>
<td>4.3.11</td>
<td>Use of letters (without FOCAL descriptors) to describe cello left</td>
<td></td>
</tr>
<tr>
<td>4.3.12</td>
<td>Use of FOCAL descriptor to represent cello left</td>
<td></td>
</tr>
<tr>
<td>4.3.13</td>
<td>All listeners, letter use within FOCAL to describe ensemble (CPVR)</td>
<td></td>
</tr>
<tr>
<td>4.3.14</td>
<td>Listener 8 depiction of ensemble (CPVR) inconsistency using letters</td>
<td></td>
</tr>
<tr>
<td>4.3.15</td>
<td>Listener 9 depiction of (Perc L) use of FOS without AREA</td>
<td></td>
</tr>
<tr>
<td>4.3.16</td>
<td>Listener 11 depiction of (Cello L) use of FOS without AREA</td>
<td></td>
</tr>
<tr>
<td>4.3.17</td>
<td>Listener 15 U-GAL depiction of ensemble (VCPW)</td>
<td></td>
</tr>
<tr>
<td>4.3.18</td>
<td>Listener 15 GAL depiction of ensemble (VCPW)</td>
<td></td>
</tr>
<tr>
<td>4.3.19</td>
<td>Overlaid descriptor plot of (Perc L) AREA descriptor</td>
<td></td>
</tr>
<tr>
<td>4.3.20</td>
<td>Overlaid descriptor of (Perc L) FOS used without AREA</td>
<td></td>
</tr>
<tr>
<td>4.3.21</td>
<td>Overlaid descriptor of (CPVN) voice described by AREA</td>
<td></td>
</tr>
<tr>
<td>4.3.22</td>
<td>Overlaid descriptor of (CPVN) voice described by FOS only</td>
<td></td>
</tr>
<tr>
<td>4.3.23</td>
<td>Descriptions favouring the FOCAL descriptor &amp; overlaid plots of FOCAL use, describing the central percussion without (figure 4.3.28) and with (figure 4.3.29) AREA descriptor</td>
<td></td>
</tr>
<tr>
<td>4.3.24 - 4.3.29</td>
<td>Differences in descriptions of the CPV narrow and VCP wide ensembles (listeners 2, 1 and 13)</td>
<td></td>
</tr>
<tr>
<td>4.3.30 - 4.3.35</td>
<td>Differences in depictions of the CPV narrow and VCP wide ensembles (listeners 11, 21 &amp; 20)</td>
<td></td>
</tr>
<tr>
<td>4.4.1 - 4.4.6</td>
<td>Ambiguities in descriptions of the solo cello stimuli</td>
<td></td>
</tr>
<tr>
<td>4.4.7</td>
<td>Listener 6 depiction of (Cello L)</td>
<td></td>
</tr>
<tr>
<td>4.4.8</td>
<td>Listener 6 depiction of (Cello C)</td>
<td></td>
</tr>
<tr>
<td>4.4.9</td>
<td>Listener 5 depiction of (Cello L)</td>
<td></td>
</tr>
<tr>
<td>4.4.10</td>
<td>Listener 19 depiction of (Cello L)</td>
<td></td>
</tr>
<tr>
<td>4.4.11</td>
<td>Listener 19 depiction of (VCPW)</td>
<td></td>
</tr>
<tr>
<td>4.4.12</td>
<td>Listener 16 depiction of (Cello L)</td>
<td></td>
</tr>
<tr>
<td>4.4.13</td>
<td>Listener 16 depiction of (VCPW)</td>
<td></td>
</tr>
<tr>
<td>4.4.14</td>
<td>Listener 8 depiction of (Cello L)</td>
<td></td>
</tr>
<tr>
<td>4.4.15</td>
<td>Listener 8 depiction of (Cello C)</td>
<td></td>
</tr>
<tr>
<td>4.4.16 - 4.4.19</td>
<td>Letter locations plotted for all listeners</td>
<td></td>
</tr>
<tr>
<td>4.6.1</td>
<td>GAL development model stage IV: Suspected ambiguities following the evaluation of U-GAL</td>
<td></td>
</tr>
<tr>
<td>5.1.1</td>
<td>U-GAL: A descriptive graphical language</td>
<td></td>
</tr>
<tr>
<td>5.1.2</td>
<td>The language development process</td>
<td></td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 2.2.1 Information about the pilot investigation programme material 60
Table 2.3.1 Programme material for descriptive graphical language investigation II 75
Table 2.4.1 Programme material for descriptive graphical language investigation III 87
Table 3.2.1 Stimuli used in the clarification of listener descriptions 113
Table 4.2.1 Stimuli used in the U-GAL investigation 151
Table 4.2.2 Listening group composition 152
LIST OF PUBLICATIONS


ACKNOWLEDGEMENTS

Many thanks to Francis Rumsey for his patience with the research (and with the researcher!) over the past years. His ever timely and much needed criticisms, witticisms and, above all, advice have shaped what was an experiential amorphous mass into the structured thesis you see before you.

Thanks also must go to the other staff and students of the Institute of Sound Recording at the University of Surrey for their constructive appraisal of the project. Russell Mason deserves a particular mention for his understanding of the on-site recording studios and for the answers to many, many queries.

For providing me with everything practical I have needed to complete the research (from cars to listeners) and his enthusiasm for the embryonic graphical language, I am ultimately grateful to Tim Nind at Harman/Becker. The employees of Harman/Becker in Bridgend have also been indispensable. To my dedicated listeners, very many thanks for your continued co-operation by participating in listening sessions and discussion groups (even though you knew “that cello” would make a reappearance). Special thanks to Rosemary Kriescher for making sure everything ran - and continues to run - smoothly, to Viv and Garry for their practical support when it came to the vehicles, and to Iraah for her work in organising listening tests. A final mention must go to Toby Newman for his adept programming of U-GALUI and for his assistance during numerous listening sessions.

It is now my turn to reciprocate a vote of thanks to Jan Berg for the verbal elicitations which have become a feature of Audio Engineering Society conventions and conferences, each one contributing something new to the research. Whilst with the AES, gratitude also to my colleagues in the discipline of Perception and Listening Tests; feedback following workshops and technical paper sessions has provided me with many opportunities for developing my thoughts. Thanks also to Neil Harris, for his open-minded discussions and continued interest in the research.

Away from audio, there are numerous people who deserve a mention. In particular, I owe thanks (and a few apologies!) to Andi for his ceaseless support, proof-reading, re-reading, advice, ability to see the whole picture and for putting up with me. And to my family and friends: Yes, maybe it is about time for me to get a "proper" job.

Finally, acknowledgement must go to Dartington College of Arts for encouraging the alternative ways of seeing that I am only now beginning to appreciate, and to my violin, a long time companion and restorer of lost reason.
CHAPTER 1

Introduction
Historical perspectives
Statement of research objectives
Evaluating the effectiveness of the research
Original contributions of the author
1.1 Introduction

Listeners will always be the final judges of reproduced sound quality. Accordingly, the involvement of listeners in the evaluation of audio devices¹ is imperative to provide an indication of the devices' sonic qualities as perceived. Essentially, by including listeners in the evaluation process, a correlation between physical and perceptual attributes is possible. However, as well as providing perceptual measures of physical qualities, subjective evaluation has the potential for eliciting novel information from listeners regarding their experiences² when auditioning reproduced audio. This more exploratory research is beneficial as it enables an understanding to be obtained regarding the salient perceptual qualities of audio devices, and consequently it is particularly advantageous in the early stages of an audio device's development.

The development of multichannel stereo and surround sound systems³ has provided the listener with an opportunity to experience enhanced spatial characteristics⁴ when auditioning reproduced audio. However, the comparatively early stage in the development of multichannel audio means that the salient qualities of listeners' auditory spatial experiences have yet to be fully explored or understood. It is therefore essential that researchers and manufacturers look to elicit descriptions of auditory spatial experiences from listeners in order to discover valid percepts for the future evaluation of spatial audio quality and the continued development of multichannel audio.

Nevertheless, eliciting descriptions from listeners is not without ambiguity. Fundamentally, experiences are independent from the way in which they are described. Specifically, at their inception, experiences are without the structure of a particular language; the role of language being to mould these pre-logical experiences into a form which can be communicated. Consequently, rather than accept a description as an inherently effective representation of an experience, it is advantageous to consider the process of getting from the unstructured experiences to their linguistic representation. Furthermore, to ensure the researcher can understand the listeners' auditory spatial experiences, it is essential that this stage in the communication process (from listeners' representation to researcher's understanding) is also investigated.

Thus, the challenge becomes one of identifying an effective communication medium; a medium which enables the listener to structure, represent and communicate their auditory spatial experiences in a manner which allows for the understanding of these auditory experiences by the researcher.

¹ An audio device is defined as any product or technology used in the recording, manipulation or reproduction of audio, or any effect of this mechanical reproduction process (for example the recording itself).
² An experience is defined as the "felt apperceptive mass to which we can inwardly point" (Gendlin (1962) 1997, p27).
³ See Rumsey (2001, pp82-118) for a comprehensive account of multichannel stereo and surround sound systems
⁴ Definitions of spatial attributes are provided later in this chapter (section 1.2.6).
It is the aim of this thesis to describe the process of developing and evaluating a descriptive graphical language suitable for the structuring, representation and communication of listeners' experiences of the spatial attributes in reproduced audio.

Section 1.4 of this opening chapter presents the research objective in more detail and displays a model which visualises the descriptive process required to fulfil this aim. Section 1.5 identifies how the effectiveness of process will be evaluated. Accordingly, this section introduces the criteria used to determine the success of the research project at each stage in the development of the graphical language. In section 1.6 a statement is provided proposing the original contributions of the author (within the context of contemporary and historical subjective spatial audio evaluation) in the development of the descriptive graphical language. An overview of the thesis is provided at the end of this opening chapter in section 1.7. Although brief, the section outlines what the reader can expect to find in the thesis and where this information can be found.

The main body of this chapter (section 1.2) concerns the context into which the research will be placed. Specifically, the section expands on the rationale for using listeners in audio evaluation; introduces suggested methods for obtaining meaningful responses from subjective evaluation; provides an alternative perspective on the role of the listener in audio evaluation; further unravels the issues involved with communication and the associated problems of obtaining meaningful responses; presents an overview of some historical and more contemporary methods used in the elicitation of responses from listeners; and finally makes the move away from using a primarily verbal descriptive medium to other descriptive (and in particular graphical) languages, concentrating primarily on the responses that have been obtained using alternative communication media. The principal arguments presented in section 1.2 are then summarised in section 1.3.
1.2. Historical perspectives – setting the research context

1.2.1 Perspectives on the use of listeners in the evaluation of reproduced audio

The involvement of listeners in the evaluation of reproduced audio is not novel. In 1980 Lipshitz and Vanderkooy suggested there to be real benefits to subjective evaluation since,

   not every audible characteristic of some components (for example, loudspeakers) can yet be objectively measured in a way which correlates meaningfully with what is heard.

   (Lipshitz and Vanderkooy, 1980, p2)

The distinction between subjective experience and objective measure was earlier acknowledged by Heyser, who wrote in the Journal of the Audio Engineering Society that “One of the worst kept secrets in audio engineering is that what we hear does not always correlate with what we measure” (Heyser, 1974, p674). Twenty five years later, a similar rationale for conducting subjective audio evaluation was offered by Moulton who stated,

   it is essential to know what ‘sounds good,’ in order to develop effective and valid correlations of the psychological quality of ‘sounds good’ with physical design qualities.

   (Moulton, 1999, p1)

Thus, audio professionals were, and continue to be in agreement about the necessity of using listeners in the evaluation of audio devices to provide a human, subjective measure of the qualities of a physical audio system. For Toole, this subjective measure was required to be universal, defined by the author as possessing “meaning outside the time and place of a specific listening experience” (Toole, 1982, p432).

But what actually constitutes meaning?

Throughout the history of subjective audio evaluation, the pursuit of meaning has been strongly correlated with a search for objectivity. For Lipshitz and Vanderkooy an individual’s preference or opinion about a device could only become meaningful to others when this had the attributes of an objective measure, and demonstrated some “elements of objectivity, consistency and repeatability” (Lipshitz and Vanderkooy, 1980, p3). Meaning in subjective audio evaluation could therefore be equated with the more objective concept of reliability; defined by Coolican (1996, p50) as the consistency or stability of a measure.

Since the listener provides the measure in subjective audio evaluation, the next challenge is ensuring the reliability of the listener. But, as Toole explained, obtaining meaningful measurements using listeners is never easy:

   People like to please and will satisfy an experimenter by producing reams of responses. Without adequate experimental procedures and controls, however, the responses tend to be rather variable. (Toole, 1982, p431)

The implementation of strict controls in listening tests is vehemently supported by Lipshitz and Vanderkooy, who argued that these controls are necessary to,
transform subjective evaluation to an objective plane so that preferences and bias can be
eliminated, in the quest for determining the accuracy of an audio component. (Lipshitz
and Vanderkooy, 1980, p1)

It is therefore pertinent at this stage to identify the nature of the controls necessary to effect this
transformation.
1.2.2 Methods for obtaining meaningful results from subjective evaluation

In his 1996 publication, Coolican listed many of the causes of unwanted variability when performing subjective evaluations. More specifically to audio evaluation, Toole's research into the subjective measurements of loudspeaker sound quality and listener performance (1985) listed a number of factors which he believed could have a bearing on the formation of a listener's judgement and thereby disrupt the reliability of any results obtained using this subjective measure. Included within Toole's list of nuisance variables were;

(i) factors of the physical environment (for example the listening room, loudspeaker position, listening position, relative and absolute loudness of the devices to be compared and the programme material).

(ii) bias caused by the investigation procedure (broadly speaking, any variability introduced as a result of the listener's participation in the evaluation).

(iii) factors associated with the listener themselves (for example, knowledge of the product under test; familiarity with stimuli, room or task; judgement and hearing ability; and any relevant accumulated experience).

A great deal has been written about controlling the various influential factors within subjective audio evaluation. The following sections serve as a reference to this body of work.

The influence of the loudspeaker and listening location in the listening room

The performance of loudspeakers in different listening environments has previously been documented by authors including Allison (1974), Salmi and Weckstrom (1982), Olive et al. (1994) and Toole and Olive (1994). Investigations undertaken by these authors identified how a loudspeaker's location within a room modified both the timbral and spatial content of an audio reproduction. Results from the preference studies of Olive et al. (1994) and Toole and Olive (1994) indicated that altering the location of a loudspeaker could have a greater influence on a listener's opinion of reproduced sound quality than altering the loudspeaker itself. Data from Toole and Olive (ibid.) further suggested that when the same loudspeaker was rated in a different listening location, a variety of preference grades could be obtained. When Zacharov investigated multichannel audio reproduction in 1998, he identified that listening location had a significant influence on a listener's judgement of envelopment and spatial naturalness.

The choice of stimuli for subjective audio evaluation

Most researchers involved in subjective audio evaluation are familiar with the importance of choosing appropriate programme items (stimuli) for use in an investigation. In his 1985 paper, Toole stated that "choosing program material represents one of the most obvious opportunities for prejudicing the results".

---

6 See chapters 2 and 3 in Coolican (1996)
6 Spatial Naturalness was defined by Zacharov (1998) as the similarity between the (non-specified) spatial qualities of the presentation and a real-life experience. Envelopment was defined in terms of whether the listener felt enveloped by the sound. Listeners were asked whether the rear sound events enveloped them completely - an effect likened to standing in the rain - or not at all (mono).

N. Ford, Doctoral Thesis, 6
of listening tests” (Toole, 1985, p6). In a later investigation, Toole and Olive (1994) found the choice of programme material to influence listeners’ preference for different loudspeakers; a result which they primarily attributed to the stimuli revealing different problems in the loudspeakers and occasionally to listeners’ tastes in music. In the preference study, four commercially available recordings were used. However, the use of existing recordings may be criticised on two counts. Firstly - as cited by Lipshitz and Vanderkooy (1980) - the researcher’s relative lack of familiarity with the signal processing or recording technique used in the creation of the stimuli and, secondly, the possibility that a listener may describe a stimulus based on their pre-existing knowledge of how this should sound. Other considerations when choosing the stimuli for subjective audio evaluation were mentioned by Rumsey (1998). He stated that stimuli should not be too complex or variable over time, since either would make them hard to grade reliably. Olive et al. (1994) stated how using short, repeatable material could increase the efficiency and ease of the listeners’ task, since this would allow the spectral and temporal characteristics of the material to become more static. Both Rumsey and Olive et al. affirmed the requirement for the chosen stimuli to reflect the demands of the particular evaluation.

The influence of loudness
Toole (1985), Bech (1998) and Aarts (1991, 1992) all acknowledged that the perceived attributes of an audio device can be influenced by differences in the relative loudness of the individual reproduction channels, suggesting loudness to be a potentially confounding variable in subjective audio evaluation. Both Bech (1998) and Aarts (1992) investigated how to objectively calibrate the relative loudness of the individual channels within a reproduction system so that they would be perceived as equal when evaluated by listeners. Results established that loudness aligned using a B-weighted pink-noise signal corresponded closely to a subjectively equalised loudness level for a pink-noise signal when reproduced over loudspeakers. Aarts (1991) also identified that using a dynamically varying signal (including extracts of classical and pop music) made no significant difference to the correlation between perceived and measured loudness. Suokuisma et al. (1998) and Zacharov et al. (1998) provide further details about the identification of appropriate signals for use in loudspeaker loudness calibration.

Bias introduced by the experimenter and experiment process
In Pitfalls in Human Research, Barber (1976) highlighted five pitfalls to be avoided when running an investigation. Consequently Barber argued the case for double-blind investigation procedures (where neither participant nor experimenter are aware of the hypothesis being tested), declaring that a participant in an investigation can be influenced by factors as diverse as the friendliness of the experimenter, to the experimenter’s desire for a particular result. Thus,

---

7 Although problematic for descriptive studies, it is believed that the knowledge of how a source should sound is beneficial when assessing the fidelity of a system. Olive (referring to Toole (1985)) stated that, without this knowledge, the task of the listener is complex since they must “form opinions on the accuracy of the test objects by listening to recordings through loudspeakers, without knowing how the recordings should sound or whether they are indeed accurate” (Olive, 1994, p1).
the only way to ensure that the 'experimenter' is unable to 'bias' the result is to prevent him from meeting the subject or to know anything of the context from which the data was collected. (Banister et al. 1994, p7).

Another cause of unwanted variability in subjective evaluation is the presentation order of items within an investigation. Order effects can be minimised by randomising the presentation order of the various factors within an investigation. Further information about order effects and how to reduce their influence is provided by Meilgaard et al. (1991, p37) and Coolican (1996, p27).

The importance of blind testing in subjective evaluation

Another important finding was that listeners in subjective audio evaluation can be affected by visual cues when assessing the sound quality of an audio device. In their 1994 study, Toole and Olive established that a listener's preference for a loudspeaker differs according to whether or not the speaker can be visually identified. The authors went so far as to state that "when listeners knew what they were listening to, the opinions were dictated more by the product identity than by the sound" (Toole and Olive, 1994, p6). Toole and Olive's results illustrate why blind testing - where the aim of the evaluation is not made known to the participant prior to their participation - is advocated for subjective evaluation8.

---

8 For more information on blind and double-blind testing see Coolican, 1996, p34.
1.2.3 Conducting dependable evaluation using human subjects

Other methods of obtaining meaningful results from subjective evaluation have been identified away from the broad discipline of audio engineering. In particular, the established area of food science can provide practical examples of how to conduct dependable subjective evaluation. Furthermore, researchers in this discipline face many of the same issues as those encountered by audio evaluators as illustrated in the following excerpt:

Only human subjects can measure sensory attributes and are, therefore, needed to correlate with and explain the physical and chemical measurements recorded with instrumentation. The detailed description of the chemical and physical sensory attributes, their precise definitions, and the exact procedures used during evaluation can often provide the starting place...[for] establishing protocols and direction for the instrumental analyses (Civille and Lawless, 1986, p213)

According to Meilgaard et al. (Meilgaard et al., 1999, p2) several factors require optimisation before “dependable sensory analysis” can be realised. Firstly, the precise definition of what will be measured; secondly, the employment of a test design that leaves “no room for subjectivity” and acknowledges sources of bias, thus minimising “the amount of testing required to produce the desired accuracy of results”; and thirdly, the training of test subjects to give reproducible verdicts. For the authors, problems in subjective evaluation are caused because people (i) vary over time, (ii) are variable amongst themselves and (iii) are prone to bias. To obtain accurate responses, Meilgaard et al. suggested that, amongst other things, repeats are made of “tasters” judgements, a large enough sample of participants is involved (between 20 and 50) and that an appropriate response scale devised.

Devising an appropriate scale for obtaining accurate responses

Subjective audio evaluation exists in many forms, from the more emotive expression of a preference for a particular product, to the judgement of the characteristics of - or similarities and differences between - audio devices. If the aim of the subjective evaluation is the latter, the participant may be provided with a set of descriptive scales on which to indicate the degree to which a particular characteristic defines a specified audio device9. When using descriptive analysis10 to obtain an “objective description of a product in terms of perceived sensory attributes” Civille and Lawless (1986, p203) noted that product descriptors - and accordingly response scales - required the use of words. Consequently, with language playing a central role in “determining the accuracy and potential benefits of a given evaluation” (ibid.), the authors believed choosing appropriate descriptive terms (to anchor the descriptive scales) to be an important issue. To obtain the most accurate responses from their tasters Civille and Lawless proposed that the terms used in the response scales should enable differentiation between products and each product to be recognised from its verbal description. Furthermore, the authors proposed that:

---

9 Exemplary response scales are provided by Toole (1985) and Gabrielson and Lindström (1985)
10 Descriptive analysis is defined by Hootman (1992, p1) as "The sensory method by which the attributes of a food or product are identified and quantified using human subjects who have been specifically trained for this purpose."
The graphical representation of listeners' auditory spatial experiences

Chapter 1 - Introduction

(i) verbal terms should be orthogonal (uncorrelated with each other)
(ii) terms should be based on the underlying structures (the texture, aroma flavour, colour etc) of the products being evaluated - as "errors in identification of natural categories often stem from ignorance or inattentiveness to attributes, and ignorance or exaggeration of the underlying structure" (ibid. p209)
(iii) a broad set of references should be employed (for example when referring to the concept of blue, more than one shade of blue should be employed as a reference)
(iv) terms should be precisely defined
(v) terms should be primary (elementary) rather than integrated (combining several terms).

In defending this last prerequisite, Civille and Lawless asserted that,

inexperienced observers in a new problem solving task will often represent stimuli holistically, rather than breaking them into their analytical parts. While such integrated terms may be useful for advertising purposes, they provide little actionable information to product formulators who need objectively anchored and elemental sensory information in order to adjust ingredients. (Civille and Lawless, 1986, p212)

Civille and Lawless maintained that the development of precise and reliable descriptive terminology would enable the taster to characterise attributes of a product and identify anomalies where objective measurements would lead to confusing results. Furthermore they believed it would help detect problems when a product did badly in a consumer test and could be useful in the construction of consumer questionnaires.

In addition to basing descriptive scales on the underlying factors of the product being evaluated, Meilgaard et al. (1999) were in support of training participants in descriptive analysis investigations. Essentially, where the aim of the subjective evaluation is for the measuring device (the taster or listener) to provide an account of the object under study, the participants are themselves responsible for some of the unwanted variability in the evaluation process. One source of variability in a participant's judgements is believed to be their previous experience.

Defining experience
Toole defined experience as "the sum total of critical listening practice, memories of reference sounds, expectations of reproduced sounds, sensitivities to various audible defects, and so on" (Toole, 1985, p7). Experience was defined by Bech (1989) as any prior (accumulative) experience with factors which could influence a listener's ability to assess reproduced sound. Amongst the factors cited by Bech were attending live concerts, the ability to play a musical instrument and experience of critically listening to either live or reproduced sound. Mattila and Zacharov made the distinction between listening experience and listener expertise. The authors stated that experience was less desirable in participants

11 Authors who have written about the influence of listener experience include Kirk (1956), Gabrielsson (1974), (1985), Bech (1989), (1992) and Toole & Olive (1994)
than expertise; defined as “definable discrimination skill with associated reliability” (Mattila and Zacharov, 2001, p2). The definition of a listener as an expert was investigated further by Shlien and Soulodre. In their investigation, the authors looked at the ability of experienced listeners to detect artefacts in digital audio. It was identified that although a listener may be sensitive to one type of artefact, they may be relatively insensitive to others. Shlien and Soulodre concluded that “the ability of an individual to act as an expert listener in a subjective test depends on the type of artefacts to be detected in that test” (Shlien and Soulodre, 1996, p10).

**Selecting listeners for subjective listening evaluation**

Since a listener’s experience affects their judgements - with experienced listeners identified (by Bech 1989 and 1992) as having an improved capacity for consistency and a greater aptitude for detecting small differences in reproduced audio - researchers have concerned themselves with devising listener selection procedures to identify those listeners best suited to participating in subjective audio evaluation. For both Bech (1992) and Mattila and Zacharov (2001) the selection process begins with applicants responding to a questionnaire about their audio history. Amongst the issues addressed by Bech were whether the prospective participants listened to music on a daily basis, if they played an instrument - and if so, how many years instruction had they had, whether (and how often) they listened to live music, if they had participated in previous listening tests, and the type of music they preferred.

Mattila and Zacharov’s Generalized Listener Selection (GLS) procedure was believed to offer a “rapid means of evaluating the suitability, discrimination ability and repeatability of an individual in a number of different audio evaluation tasks” (Mattila and Zacharov, 2001, p2). The GLS procedure consisted of three distinct stages: (i) a questionnaire; (ii) an assessment of applicants’ hearing acuity and; (iii), listening tests to establish applicants’ reliability and discriminatory ability. Although the authors believed the selection process could identify those listeners best able to reliably discriminate between audio devices they also asserted that - since only highly sensitive listeners were chosen for the listening panel - caution should be taken when attempting to infer results from this panel to the opinions of a more general public.

**Benefits of experience and expertise**

In his 1992 investigation, Bech compared responses from professional listeners with those from less experienced listeners. His results indicated that experienced listeners were in more agreement about their judgements than inexperienced listeners. This difference in consistency allowed Bech to calculate that fewer experienced listeners (one compared with up to seven inexperienced listeners) were required to obtain the same statistical result. Bech concluded, “thus there appears to be quite an advantage in using groups of subjects with a high level of preexperience in critical listening” (Bech, 1992, p594).

Gabrielsson et al. (1974) asked three groups of listeners (listeners in general, musicians and hi-fi listeners) to evaluate the qualities of various loudspeaker reproductions. Results identified that no significant differences existed between the listening groups, but the least experienced listeners used
higher fidelity ratings than the remaining listeners. When listeners were required to state their preference for a particular loudspeaker in Toole and Olive's 1994 investigation, similar results were found. In this study, listeners who had previously participated in controlled listening tests used lower ratings than their less experienced counterparts. It was of further interest to observe that both groups of listeners responded similarly when they could visually identify the loudspeakers; further justifying the preference for blind testing in subjective audio evaluation. In another investigation, Gabrielson et al. (1974), found there to be a significant interaction between the type of reproduction - which differed in terms of loudness and frequency - and the listeners' experience; either as a hi-fi listener or a listener in general. When summarising the results of both investigations, they stated that listeners in general were less reliable than either the hi-fi listeners or the musicians and tended to give higher ratings to the poorest reproductions.

In a later experiment (Gabrielsson and Lindström, 1985) listeners were once again divided into three groups: (i) hi-fi experienced (ii) musicians and (iii) inexperienced. It was established that experienced listeners provided the most reliable ratings and differentiated more clearly between the loudspeakers under test. It was further determined that experienced listeners and musicians used the various rating scales similarly. However, the inexperienced listeners provided inconsistent ratings and preferred the system considered poorest by the more experienced listeners. In an earlier study of 210 students by Kirk (1956), a listener's prior auditory experience was also found to influence their preference for reproduced audio, with more advanced musicians responding differently to those listeners who were comparatively new to musical study. Musicians in the investigation preferred stimuli which matched their existing knowledge of how the material should sound; an intuitive result since a person can only ever hear in the context of what they currently know. Kirk's research was cited by Gabrielsson and Lindström as a plausible explanation for the least experienced listeners in their experiment preferring the cheapest system. Kirk went on to manipulate the current knowledge of his listeners and in doing so, altered their preference for different audio reproductions.

Training listeners
According to Meilgaard et al., when responses to the same object show there to be differences between individuals, sensory analysts should be aware that this could result from differences in

the sensation [participants] receive because their sense organs differ in sensitivity or by a difference in their mental treatment of the sensation e.g., because of a lack of knowledge... or because of lack of training in expressing what they sense in words and numbers. Through training and the use of references we can attempt to shape the mental process so that subjects move toward showing the same response to a given stimulus.

(Meilgaard et al., 1999, p3)

Training is thus acknowledged as a means of manipulating a listener's experience and acuity to achieve greater discrimination between objects and more reliable results¹². In their 1980 paper, Lipshitz and

¹² The benefits of training participants is covered by Meilgaard et al. (1999), Olive (1994) and Bech (1990, 1992)
Vanderkooy stated that using trained listeners increases a test's sensitivity. Accordingly, the authors declared that "only trained, experienced listeners should be part of the listening panel if greatest sensitivity is desired." (Lipshitz and Vanderkooy, 1980, p7). Bech contended that training is "an attempt to eliminate or control a number of variables that will otherwise have an influence on the reported results" (Bech, 1990, p111). In a later paper, Bech maintained that even listeners who commence an experiment consistently able to differentiate between objects, need to "familiarize themselves with the task" by participating in training investigations (Bech, 1992, p590). For Bech, the aim of training was to ensure that listeners had reached a point where their performance in listening tests was constant over time, indicating that the learning process had stopped. Bech determined that 65% of listeners would reach stability after four training investigations. Another statistic of interest was reported by Olive. He suggested that, since trained listeners are "more critical and consistent in their opinions" (Olive, 1994, p1), they can produce more statistically reliable data than untrained listeners. Olive demonstrated that the ability of a listener improves 23% after five training sessions. In a more recent 18-month study, Olive (2003) compared the abilities of 256 untrained listeners with 12 trained listeners when stating a preference for different loudspeakers. Although preference was comparable between listening groups, trained listeners were found to be better able to discriminate between different loudspeakers and provided more repeatable judgements than their non-trained counterparts13.

In his 1994 study, Olive advocated the use of a computer-automated training tool to improve the listener's ability to recognise, describe and reliably rate timbral differences on a 10-point preference scale. Olive suggested that the results of his training programme could be used as a basis for selecting the best listeners to participate in subjective audio evaluation. Shively and House (1998) studied listener training to establish a method for improving listeners' repeatability in automotive audio evaluation. Like Olive, the authors used a self-administered computerised training tool and supported a two-phase training method. The first phase was to educate listeners about the identification of spectral peaks and dips in programme material, whilst the second trained listeners in the use of preference ranking and timbral balance14 scales, ensuring that the first part of the training was being applied correctly. Listeners trained using Shively and House' method were subsequently assessed to ascertain the number of trials required before they could repeatedly rate to within 0.5 points on a 10-point scale. It was established that stability was reached after 5 tests.

**Concluding remarks**

Where the aim of subjective evaluation is to obtain from the listener a reliable (objective and thus meaningful) measure of the perceived qualities of an object under study, the rigorous control of numerous factors is advised throughout the reviewed literature. Physical factors (such as the listening location or loudness of the stimuli) can be equalised or included in the investigation as factors of

---

13 It was noted by Olive, that listening position could have introduced a confounding factor into his study. Trained listeners participated individually in the investigation and sat on-axis in relation to the loudspeakers. The 256 untrained listeners participated in groups of eight and were randomly assigned a seat from two rows of four.

14 Timbral balance scales divide frequency into three bands (treble, midrange and bass). Listeners were trained by Shively and House to associate timbral changes with the appropriate frequency ranges.
interest, whilst the randomisation of stimuli, and the use of (double) blind methods, can compensate for the unwanted affects of variability introduced by the investigation itself. Finally the test methodology and listener can themselves be manipulated, with the response method selected for its descriptive accuracy and the listener selected on the basis of their prior knowledge or expertise, and trained to improve the reliability (the objectivity) of their measurements and the correctness of their responses.
1.2.4 An alternative perspective on the measurement of meaning

As demonstrated by the review of available literature, meaning has traditionally been ascribed to a measure of the external object (the audio device under examination as perceived by the experimenter) and not to a measure of the listener's perception; meaningful results occurring when the listener provides a reliable measure of what is already known about the physical properties of the object. It is this object which provides the benchmark for evaluating accuracy, and this object against which the listener's responses are validated. The higher the correlation between the listener's response and the object's known characteristics, the greater the accuracy and meaning associated with the results.

Although rigorously controlling factors in subjective audio evaluation provides a means of acquiring a highly reliable measure of the audio device under scrutiny, there is a risk that the obtained results will bear little relation to the device as perceived by the listener.

The problem of neglecting an individual's perception

The opportunity for the researcher to overlook the listener's actual perception is alluded to in the following statement by Toole:

Humans, in most real-world situations, are notoriously unreliable as judges not because they are inherently capricious, but because they are sensitive to a host of factors other than the one that may be of particular interest (Toole, 1982, p431).

The presence of an additional "host of factors" alongside those being examined, suggests that researchers may not always measure the most salient percepts of the participants in their investigation. Instead the researcher may choose to focus on a dimension which, although objectively measurable, is not found to be perceptually relevant. Since "you only get an answer to what you ask" (Kjeldsen, 1998), it follows that a listener's true perception could remain hidden if it is not the focus of the investigation. According to Köster, asking inappropriate questions is a problem for human research because "whenever one puts a question to a subject, one always gets an answer, even if the question is in principle unanswerable." Köster continues, "of course there is no harm in asking useless questions, as long as one does not base conclusions on the answers" (Köster, 2003, p364).

To avoid this possibility, Guski believes that, rather than assume they know the possible responses of their listeners, researchers should - as least as a first step - use more exploratory methods to obtain "free descriptions of the sound, unconstrained by suggestions or by a certain response format" (Guski, 1997, p768). Toole notes that allowing listeners to provide their own comments (alongside any required responses) provides a useful means of correlating the questions asked in an investigation with the listener's actual percepts, as "we are still learning what questions should be asked" (Toole, 1985, p30).

Once the relationship between physical properties and perceptual attributes has been defined, Guski contends that it is possible to move on to undertaking more psychophysical, confirmatory studies,
getting listeners to judge audio attributes using response scales. When considering judgements about sound quality, Blauert and Jekosch\textsuperscript{15} believe that measures should avoid using psychoacoustic techniques that disregard any meaning a listener may associate with the auditory event. Rather, sound quality evaluation should be regarded as more psychological in nature since “the output of auditory perception is not at all predetermined solely by the acoustic input to the auditory system, but results from a complex interaction of auditory input, non-auditory input, expectation and mood” (Blauert and Jekosch, 1997, p751). When measures are more psychological an important task of the researcher is to “select representative listeners for evaluation procedures” however Blauert and Jekosch also note that these representative listeners are “not necessarily expert listeners.” (ibid.)

A counter-argument for listener selection and training
Selecting and training listeners for participation in subjective audio evaluation provides the researcher with obvious opportunities for biasing an investigation in their favour. Rejecting listeners at the outset of an investigation, excluding them following training, or omitting their results from any analysis - because of a lack-of-fit with other participant’s data or a conflict with existing assumptions about what constitutes an accurate measure of an audio device - can result in conclusions which only fulfil the researcher’s prophecy of what should happen in an investigation.

For Meilgaard et al., training the listener increases their sensitivity to the factors under investigation and puts them “in a frame of mind to understand the characteristics” the researcher wishes them to evaluate (Meilgaard et al., 1999, p37). However, not only can this manipulation once again result in findings which neglect the listener’s actual perception (if these are different from the researcher’s), but training influences how the listener will react to an audio device on subsequent meetings, resulting in accurate responses for the researcher, but questionable validity for the listener. Although training appears appropriate when the qualities to be measured are known, as Berg declares; investigations which employ highly trained listeners “may suffer, especially in relatively unexplored areas of subjective judgement, from the danger of ‘training out’ real and important differences between subjects” (Berg, 2002, p3, publication 1). Berg continues “it is possible that using such rigorous training one might end up getting the answer the subjects were trained to provide, rather than that which they might have provided if left more to their own devices” (ibid.). Thus more experienced listeners - although able to provide consistent, uniform responses - bring with them the ‘baggage’ of their previous listening tests and any training they may have received, and this may prevent the researcher from uncovering a listener’s actual experiences.

When offering an overview of Personal Construct Theory, Bannister contended that traditional investigations describe ‘real’ situations from the point of view of the experimenter, “the situation as the

\textsuperscript{15} Sound quality was defined by the authors as "a descriptor of the adequacy of the sound attached to a product. It results from judgements upon the totality of auditory characteristics of the said sound - the judgements being performed with reference to the set of those desired features of the product which are apparent to the users in their actual cognitive, actional and emotional situation" (Blauert and Jekosch, 1997, p750)
The graphical representation of listeners' auditory spatial experiences

Chapter 1 - Introduction

The experimenter conceptualises it. Consequently, "responses which are not meaningful in terms of this view are regarded as without content. The subject failed to learn, the subject was unmotivated, the subject did not perceive a, b, or c" (Bannister, 1962, p114). Similarly, Kelly - the author behind the psychology of personal constructs - stated that, rather than ascribe failure to a participant in an investigation, it should be acknowledged that they simply did not complete the investigation in the way anticipated by the experimenter. He continued,

let us put the burden of discovery on the experimenter rather than on the subject. Let the experimenter find out what the subject is thinking about, rather than asking the subject to find out what the experimenter is thinking about. (Kelly, 1963, p77)

Forty years later, Köster produced a paper grappling with more of the peculiarities of human research. The first fallacy to be considered by the author was that of participant comparability; the belief that participants should perform the same task in the same way with any variability between them stemming from their ability rather than individual differences. According to Köster, although much effort has been made to divide subjects according to their responses or develop new scaling methods which reduce variability "there is usually little attention paid to the underlying factors that determine the individual differences. Do people perceive the stimuli differently or do they perceive the task differently?" (Köster, 2003 p360). The second fallacy to be examined by Köster was that of consistency; the (mistaken) belief that people did not change over time.

It is assumed silently that an experiment when repeated under exactly the same circumstances will produce the same results...In all of this, it is forgotten that the subject really might change, because he or she has a memory and therefore the second encounter with a stimulus may not mean the same to him or her as the first one. (Köster, 2003, p361)

Historically then, meaning has been positioned with the object and not the subject of the subjective evaluation. In their work on Qualitative Methods in Psychology, Banister et al. noted the curious anomaly associated with the terms object and subject when used in much quantitative psychology; where individuals are called subjects but treated as objects, "and we pretend to be objective but are still always deeply subjective" (Banister et al., 1994, p5). Accordingly, what appears to have been taking place in much audio evaluation is not subjective, but a form of objective evaluation using a subjective measure, where the aim of the evaluation is to provide a measure of the object under study, and the stringent control of potential sources of variability leads to accurate (valid for the object under study) results. However, it should be argued that these results can possess poor validity (not measuring what was intended) when considered with respect to their ability for providing an accurate measure of the listener's perception.

Rather than strive for objectivity in subjective research, Banister et al. assert that

It is in the nature of human beings, and a 'power' they have, to reflect upon their actions and to give account of those actions, and this means that a properly scientific approach to the study of action and experience should employ methods which engage with rather than try to screen out these powers (Banister et al., 1994, p9).
Furthermore, according to the authors, knowledge should be accepted as constructed, "as one version of reality, a representation rather than a reproduction. It is understanding in process, which is open to multiple interpretations." (ibid., p143). Thus, even though participants in subjective audio evaluation may change their responses over time, or may respond differently to one another or the experimenter, response diversity should not necessarily be restricted. Allowing listeners freedom in their response enables an insight to be gained into alternative perceptions and a valid knowledge and understanding to be obtained by illuminating the individual’s experiences of a phenomenon. As Banister et al. affirm, this provides us with “an attempt to capture the sense that lies within, and that structures what we say about what we do” (ibid., p3). The following section provides an overview of work concerned with establishing the individual as the true originator of meaning.

**Positioning meaning with the subject rather than the object of subjective evaluation**

As human beings, we attempt to make sense of all our experiences. Through our mental acts, we strive to impose meaning upon the world. (Spinelli, 1989, p1)

When Kelly developed the *psychology of personal constructs*¹⁶, the underlying philosophy was that ‘reality’ only exists as construed by each individual. According to Kelly, “Man looks at his world through transparent patterns or templets (sic) which he creates and then attempts to fit over the realities of which the world is composed” (Kelly, 1963, p9). Without their individual patterns, Kelly believed that people would not be able to make any sense out of the undifferentiated world. When Bannister offered a critique of Kelly’s theory, he wrote that items assume their meaning as a result of this creation of patterns. Hence, “the substance which he construes does not produce the structure, the person does” (Bannister, 1962, p115). In developing his theory, Kelly believed this creative process emphasised the capacity of the individual to represent rather than respond to their environment. Accordingly, each individual can place alternative constructions upon it and, indeed, do something about it if it doesn’t suit him. To the living creature, then, the universe is real, but it is not inexorable unless he chooses to construe it that way. (Kelly, 1963, p8)

In the psychology of personal constructs, each construct consists of a unique similarity/difference dimension with meaning obtained through the consideration of both poles. Tindall provided an example using the construct ‘friendly’, declaring “we would recognize that two people using the construct ‘friendly’ are experiencing different realities if we also know that the difference pole for one is ‘not so friendly’ and for the other is ‘hostile’.” She later confirmed that “it is our construction, the meanings are inferred by us. These meanings are not part of the event, not statements of reality.” (Banister et al., 1994, p74)

A second philosophy dedicated to the view of reality as experienced by - and meaning as inferred by - the individual is phenomenology. According to Köster, in phenomenology, consciousness is viewed

---

¹⁶ Personal constructs are defined by Landfield as "the system of dichotomous contrasts employed by the individual as he tries to make sense of his experience" (Landfield, 1968, p135)
not as a receptive box that is filled with images from the outside world, but as an active principle that is always directed outwards at its object and that in doing so, by its 'intentionality', provides 'meaning' to the objects it is directed at. (Köster, 2003, p368)

Giorgi presented a description of phenomenological understanding where intuitions or objects exist "in terms of the meaning that the phenomena have for the experiencing subjects" (Giorgi, 1997, p237). He continued with a visual example of how two people may view the same painting in different ways; the first person calling the painting 'ugly' and the second 'beautiful'. Thus for person one, the painting will have all the properties of ugliness and for the second, all those of beauty. However, "no claim is made that the painting is in itself either ugly or beautiful; only its presence for the experiencer counts, and an accurate description of the presence is the phenomenon" (ibid.). Thus, reality only exists as it is known by (or given to) the individual, who only knows the givens they are aware of.

In positioning 'meaning' with an external reality, traditional subjective audio evaluation may be considered anomalous. Indeed, as Cherry wrote in On Human Communication:

"We speak of the “real world” as being “outside us," a curious phrase — outside what exactly? Strictly, this is putting the cart before the horse, for if anything is “real” to each one of us, it is our experiences, our sensations. (Cherry, 1966, p263)

The rejection of an ultimate reality also negates the possibility that only one correct interpretation of an event can occur in any given situation. Spinelli (1989) states that what most of us term a ‘correct interpretation’ is not based upon objective laws or universally accepted truths, rather it is influenced by the viewpoint of a cultural consensus. Consequently, although (with adequate control) the listener can be trained to confirm subjectively what is objectively known about an audio device, it is advantageous to consider an alternative perspective on the role of the listener in subjective evaluation: A perspective which acknowledges and attempts to understand in more detail the listener’s experiences: A perspective which places the meaning of subjective evaluation back with the subject of that evaluation - the listener - and evaluates the accuracy of the subjective evaluation through its ability to measure what is known of the listener’s experiences. For, as Moustakas asserts, “only what we know from internal perception can be counted on as a basis for scientific knowledge” (1994, p45).

In positioning meaning with the listener, subjective audio evaluation is provided with an alternative perspective from which it can acquire knowledge. From this perspective, rather than attempting to answer a pre-determined question (one reflective of the researcher’s and not the listener’s experiences) subjective evaluation can explore and further understand a phenomenon through the listener’s experiences. In doing so, the research removes meaning from the external object and places it with the listener, accepting that each listener will have their own reality, their own experience of the physical attributes of the object. But how can the success of such a subjective evaluation method be measured?
Practical considerations for a more qualitative understanding of listeners' experiences

“Qualitative analyses can be evocative, illuminating, masterful – and wrong.”

(Miles and Huberman, 1994, p262)

Assessing listeners’ experiences is complex, since the phenomenon being measured is intrinsically part of the participant. Amongst the measures proposed by Miles and Huberman for verifying quality in qualitative analysis include:

(i) an assessment of the research population’s representativeness, with an acknowledgement of any implications for the subsequent generalisation of research findings to a wider population
(ii) an investigation of surprising results
(iii) a check on the meaning of outliers or extreme results (an examination of which tests the suitability of the research findings and enables the evolution of a better explanation)
(iv) the negotiating of findings with participants: “An alert and observant actor in the setting is bound to know more than the researcher ever will about the realities under investigation...In that sense, local informants can act as judges, evaluating the major findings of a study” (ibid. p275).

Ultimately Miles and Huberman commented on the importance of acknowledging the researcher's involvement in the study and any subsequent bias this may introduce into the results. When understanding any text, Gadamer observed the importance of recognising the present situation within which the text is being read by a researcher. Gadamer believed that the (necessarily) current perspective of the reader undoubtedly influences their interpretation of the work and that it is impossible for a reader to leave their immediate situation without adopting an attitude. Accordingly Gadamer suggested that rather than consider a reader's understanding of a text as a reconstruction of the text as presented by the author, it should be considered more of a mediation; with both text and reader "fused into a common view of the subject matter" (Gadamer, 1976, xix).

The researcher’s central position in qualitative research and its consequent influence on any conclusions was identified by Banister et al. as an outcome of the work’s existence in a complex world. “It involves researchers’ active engagement with participants and acknowledges that understanding is constructed and that multiple realities exist” (Banister et al. 1994, p142). Furthermore, “the ways in which we theorize a problem will affect the ways we examine it, and the ways we explore a problem will affect the explanation we give” (ibid. p13). As with Miles and Huberman, Banister et al. suggested obtaining feedback from participants when interpreting the provided data, declaring “there are no techniques or analytic procedures that escape the dangers of exploitation” (ibid. p67).

Since the inception of personal construct theory, the repertory grid technique (RGT) has been employed to understand more about a person’s constructs. For Tindall, “the aim of the repertory grid is to illuminate a person’s current understanding of whatever it is they are concerned with” (Banister et al., 1994, p75). For Kjeldsen (1998), RGT “brings into focus the individuals reflections,...thus opening a field of investigations into not only immediate responses, but the reasoning that makes people
respond as they do”. In RGT, information about how an individual understands their world is typically elicited by asking them to describe similarities and differences between three different “elements” (anything from people to loudspeakers), specifying similarities between two of the elements and difference from a third. According to Tindall, the way each participant views their world emerges throughout the process, with any understanding thus “framed by the participant.” (Banister et al., 1994, p77). Although any results from RGT may be framed by the participant, the reflexive nature of personal construct psychology dictates a central role for the researcher. Consequently Tindall states that it is once again necessary to consider the researcher’s influence; in essence, how could their particular interests, skills and understanding have affected the process?

For Giorgi (1997), the use of phenomenology in the context of psychological research involved a five-step process of data collection and analysis: (i) the collection of verbal data in terms of a faithful to the experience description of an individual’s experiences; (ii) a global reading of the data; (iii) a slower “discovery-oriented” re-reading (sufficiently open to let unexpected meanings emerge) and division of the data into identifiable “meaning units” (as defined by the researcher); (iv) the organisation and expression of the data in the disciplinary language of the researcher and; (v) the synthesis or summary of the data for purposes of communication to the scholarly community, identifying the meaning units essential for the phenomenon under study.17

In analysing phenomena, Giorgi makes the distinction between unwanted interpretation and desired description, contending that the former is distinct from the latter because, whereas description “is the clarification [in language] of the meaning of the objects of experience precisely as experienced” (Giorgi, 1992, p122), “in order to account for a phenomenon [interpretation] brings a perspective to the given, either from theory or for pragmatic reasons, that is not necessarily demanded by the intuitive evidence” (Giorgi, 1997, p241). To counter the claim that it may be necessary to interpret the data in order to account for any incompleteness or ambiguity, the author asserts: “There is no rule that data must be aesthetically elegant or logically consistent. Whatever shows up is described precisely as it shows itself” (Giorgi, 1992, p127). Thus, according to Giorgi, “with the help of free imaginative variation one describes the essential structure of the concrete, lived experience from the perspective of the discipline” (Giorgi, 1997, p247). When describing experience, Giorgi asserts that this description should not be claimed as total or complete, but that it effectively comprehends the phenomenon being presented to consciousness, and that there is a difference between accounting for a presence that gives itself as doubtful, uncertain, or plausible and one that presents itself as effective and fulfilling. (Giorgi, 1992, p122)

Through the representation and illumination of people’s experiences, Banister et al. focus on the aims of qualitative research as acquiring valid knowledge, a definition of which is given as the “adequacy of the researcher to understand and represent people’s meanings” (Banister et al., 1994, p143). For Moustakas, “In accordance with phenomenological principles, scientific investigation is valid when the

---

17 Moustakas (1994) provides a thorough account of analysis in phenomenological research in the latter part of his work.
knowledge sought is arrived at through descriptions that make possible an understanding of the meanings and essences of experience" (Moustakas, 1994, p84).

Thus, to assess the validity of the research, the researcher moves away from correlating research conclusions with an external reference. Instead, the researcher’s aim becomes one of attempting to elicit from the listener a full description of what it is they have perceived. And it is this experience that provides the initially valid premise on which conclusions can be founded. If this more phenomenological perspective is observed, it becomes necessary for the researcher to regard the process of subjective evaluation from the opposite direction to the historical perspective presented; essentially from the inside outwards. Here, the only objective certainty is the subjective experience of the individual who is tasked with listening to the reproduced sound.

The significance of a subjective audio evaluation can therefore be measured according to the ability of the listener to communicate their experiences of what they have heard; the listener’s experiences, and not an external reference, providing the meaning when the aim of subjective audio evaluation is to understand more about the experiences of a listener. However, the process of obtaining this meaningful information from the listener is not straightforward, as will be demonstrated in the following section.
1.2.5 Communicating experiences

Experiencing is *concrete*. It is the felt apperceptive mass to which we can inwardly point. It is a "this" or a "this way I feel." It is not to be equated with logical definitions and schemes, for these are abstract only. They *represent* something, but are not *in themselves* something. (Gendlin, (1962) 1997, p27)

Saussure concurred with Gendlin when he described "thought" as a "vague shapeless mass" or a "swirling cloud, where no shape is intrinsically determinate". He continued, "were it not for signs, we should be incapable of differentiating any two ideas in a clear and constant way...nothing is distinct, before the introduction of linguistic structure" (Saussure, 1983, p110). But, as Cherry declares, what is communicated with a sign is not the actual *thought*, but rather "a representation for carrying out this function under the severe discipline of using the only materials he [the communicator] has...Speech is like painting, a representation made out of given materials - sound or paint" (Cherry, 1966, p74). Thus a sign, or language of signs - be it speech, writing, drawing, semaphore, Morse code, or a chain of binary numbers - is necessary to structure, differentiate between and represent experiences in order that they may be communicated. There exists, however, a notable distinction between experience and language, a distinction captured by Gibson in his work on the *Perception of the Visual World* (1950).

Gibson's work chronicled a series of experiments which had previously been performed with patients who had recently had their sight restored. In these experiments, the patients were essentially asked to describe what they were seeing. But, the patients' response to this question was bewilderment.

The patients all had the use of language but they found it difficult or impossible to describe what they saw or to apply words to it. The question "are things projected in space?" simply did not mean anything to them.... the patient could not assign to his impressions terms like black and white, moving or still, far or near.... He could not, in fact, say anything about his visual impressions. He had these terms in his vocabulary but they referred to tactual and muscular impressions only. (Gibson, 1950, pp217-218)

Gibson attributed the patients' behaviour to the fact that people have to learn how to see the world. This process initially involves learning how to identify and discriminate between objects and only later, how to recognise and describe the qualities of, or differences between, these objects using the facility of words. In *his Cours de Linguistique Générale* 18, Saussure introduced the linguistic process as being one of inter-mediating between 'chaotic thought' and 'sound': At the heart of the process, "what takes place, is a somewhat mysterious process by which ‘thought-sound’ evolves ... and a language takes shape with its linguistic units in between those two amorphous masses" (Saussure, 1983, p111). It is useful therefore to identify these "linguistic units" and how they are related to pre-linguistic experience.

---

The definition of a linguistic sign

Rather than as a link between a thing and a name, Saussure defines each linguistic unit - or sign - as the combination of a comparatively abstract concept (the signification) with a signal or sound pattern. The signal is, in essence, a “representation of sensory impression” such as the consideration of linguistic activity or the process of talking silently to yourself. The signals most often used for communication in subjective audio evaluation are words. According to Ogden and Richards (1966), these verbal signs can be symbolic (whereby items are identified, catalogued or related to one another) or emotive, where signs are selected to elicit certain responses in the reader or listener. Crucially for Silverman, signs derive meaning as a result of their difference from other signs “so the colour red is only something which is not green, blue orange etc” (Silverman, 1993, p72).

Regardless of their categorisation, the sign, as recognised by Saussure, is innately arbitrary. More specifically, the link between signal and signification is arbitrary, with rules - rather than any intrinsic value - affording the signal its character. In his review On Human Communication, Cherry (1966) stated that words are simply signs that have achieved significance by convention. Those who are unaware of, or fail to adopt, the convention simply fail to communicate effectively. But, even when the convention is adopted, there is an inherent vagueness to signs. This vagueness led Ogden and Richards to claim that the completeness of any reference should be regarded in terms of its ability to grasp its referent, since words mean nothing by themselves (Ogden and Richards, 1966, p10). As early as 1689, John Locke, in his Essay Concerning Human Understanding, wrote: “When we begin to fix by means of words... abstract ideas... there is a danger of error. Words should not be treated as adequate pictures of things; they are merely arbitrary signs for certain ideas - chosen by historical accident and liable to change” (Locke (1689), cited Cherry, 1966, p70).

The arbitrary nature of signs

Thus, the arbitrary nature of language means that any word can be invented and used to describe an object, and any referent could just as easily be known by some other name. Take, for example, the object known as a ‘pen’ in the English language. Depending upon cultural and linguistic boundaries, the object has many different names, yet the object (either a female swan, small animal enclosure or writing implement) remains constant in its appearance and purpose. Emotive language was believed by Cherry to be an even more ambiguous communication device than symbolic language, since words which act solely as “emotive stimulants” do not name objects with precision. He provided as examples words such as ‘democracy’, ‘happiness’ and ‘civilisation’, which may be interpreted differently by each individual depending upon their history. Spinelli (1989) also positioned personal history central to an individual’s use and understanding of language. When responding to the question “what would we perceive a simple object to be if the sign we know for that object was removed?” Spinelli suggested that the object would still be something, but the definition or meaning given to that something would have as much to do with the individual, and the meaning system employed by the individual, as it would have to do with the object itself.
Chapter 1 - Introduction

The context dependency of language in communication
Not only should language be understood according to the unique history of those who use it, but - when used in communication - the meaning of a sign should also be considered in context. Cherry believes that the full meaning of a word does not appear until it is placed in context; the word’s meaning altering according to the relationship and communication experience of the communicators and the situation in which the communication is presented. As summarised in Gadamer’s discourse, “language is not the possession of one partner or the other, but the medium of understanding that lies between them...[with] the meaning of words [depending] finally on the concrete circumstances into which they are spoken” (Gadamer, 1976, xxxi). Olson and Bialystok (1983) likewise contended that the description assigned to a particular event is dependent (not only upon its particular characteristics) but the context in which the event occurs and the prior knowledge of the perceiver. Furthermore, for Cherry, the representation as presented by the communicator will only have meaning for the receiver if it represents a continuity of their own experience. Cherry states, “we may never fully understand if we are not bred in the culture and society that has moulded and shaped the language” (Cherry, 1966, p74).

Consequently for Levy, the interpretation of any event never occurs in a vacuum. “The event being interpreted is part of a context that consists of other events upon which some interpretation has already been placed, the individuals involved, who have already been characterised in some fashion, and the particular language system or theory of the interpreter” (Levy, 1963, p90). The complexity and ephemeral nature of interpreting a communicated event, whatever the medium, was summarised by Gadamer, who contended

The text or art work lives in its presentations...The variety of performances or interpretations are not simply subjective variations of a meaning locked in subjectivity but belong instead to the ontological possibility of the work. Thus there is no canonical interpretation of a text or art work; rather, they stand open to ever new comprehensions.
(Gadamer, 1976, xxvi)

Language needs therefore to be understood both as a variable within the creative context of a communication and as a variable in itself; dependent as it is on arbitrary signs. Yet, in employing this vague, context dependent, personalised language, we attempt to make sense of our experiences; structuring and differentiating between objects of experience in order that they may be represented in a (hopefully) meaningful set of terms which effectively communicate what we have experienced. As Cherry wrote: “The only way to pin down a thought before it can slip away and fly out of the window is to jump on it with both verbal feet, to pin it down with language, by diagrams, or by mathematical symbols”. Though, Cherry continues “such language may be inadequate” (Cherry, 1966, p79).

The idiosyncrasies and inadequacies of language
Much linguistic inadequacy occurs because of a language’s inability to represent the subtleties of thought. A useful analogy was provided by Gramont, who, in his introduction to Language and the Distortion of Meaning, likens the shift from thought to linguistic representation, with a shift from an analogue information system to a digital system (Gramont, 1990, p10). Briefly, not enough words are
available to express all experiences. Consequently, as Levy (1963) asserted, something is always lost when words are used to describe events. Similarly, Cytowic (1993) was of the opinion that not everything we do or know can be expressed in language. Köster acknowledged that, although people enjoy eating and drinking, they are rarely asked to analyse their experiences. Consequently, "when they occasionally do analyse them, they have great difficulty finding the right words to express their experiences" (Köster, 2003, p365). Köster later continued, "true meaning is based on very intricate and often hidden motives which, if they can be brought to consciousness at all, are difficult to express verbally" (ibid. p369). Remaining with food and its description, Civille and Lawless remarked upon a known occurrence in sensory evaluation, namely the “tip of the nose” phenomenon. Here, the individual smells a familiar fragrance, but finds “an annoying mnemonic gap from the sensation to the name” (1986, p206). Language’s insufficiencies in this respect can be summarised by Kelly:

A person is not necessarily articulate about the constructions he places upon his world. Some of his constructions are not symbolized by words; he can express them only in pantomime. Even the elements which are construed may have no verbal handles by which they can be manipulated and the person finds himself responding to them with speechless impulse. (Kelly, 1963, p16)

Kelly (ibid.) further identified that when language is used to promote - rather than to describe - an experience, the language used makes us sensitive to certain events and not to others; language therefore moulding our ways of thinking and dealing with events. Along similar lines, Novitz (1977) stated that “one’s ability to describe an object, and more particularly the way in which one describes it, often affects one’s ability to recall it.” When Schooler et al. (1993) conducted numerous experiments into people’s ability to communicate their insights, they found that verbalisation (converting thought into spoken word) occasionally impaired participants’ ability to solve problems. The authors explained that their results demonstrated how verbalisation can cause non-reportable aspects of a task to become overshadowed by those that are more readily reported.

With any language, the key criterion for Gendlin is that it does not replace the felt (pre-logical) experience of the individual. “If we do not have the felt meaning of the concept, we haven’t got the concept at all - only a verbal noise” (Gendlin, 1997, p5).

**Personal meaning in language**

Not only are thoughts difficult to express in language, but the signs eventually used by an individual are likely to be applied to their experiences with relation to their own personal meaning. Kelly therefore believes that even though each sign will possess meaning for the individual, the meaning associated with this terminology may not be universally held and consequently the terminology may represent an alternative experience in another. As Kelly asserted, “since constructs are primarily personal, not all of them are easily shared. The particular nature of a person’s construct or his unusual use of terminology may be misleading to his listener” (Kelly, 1963, p116).
In an investigation setting, the individuality of personal language is potentially problematic. As Rumsey contended it is "quite likely that each individual using these terms\(^{19}\) will think that everyone else understands the same thing by them, but the literature is full of subtly different interpretations" (Rumsey, 2002, p661). As Letowski surmised, the number of terms used when describing a sound "is a blessing for artistic freedom, but it is a problem when it comes to meaningful communication between people" (Letowski, 1989, p2). And for Toole, "the language of critical listeners tends to be closer to the language of poetry than of scientific measurement" (Toole, 1985, p9). Accordingly it is not possible to say with certainty that each individual within an investigation is using the same language in the same way as each other or as intended by the experimenter. Ogden and Richards similarly reasoned that participants in conversation will, normally, spring spontaneously to the conclusion that "the speaker is referring to what we should be referring to were we speaking the words ourselves" (Ogden and Richards, 1966, p15); a conclusion that can be misleading. 

The individuality of language was borne out by the findings of Bannister (1962). In his experiment, participants were asked to rate 20 photographs of different people in terms of seven adjectives which were not defined for them\(^{20}\). Results identified that, although participants agreed when rating on the different adjective dimensions (applying universal definitions to adjectives), there was little agreement in how the participants rated the different photographs using the same adjectives. It was apparent for Bannister that, even though individuals were assigning common meanings to the adjectives, these adjectives were being applied differently. Thus Bannister concluded that it was more meaningful to describe construct systems as independent of the particular elements construed. An overview of the ambiguity which can occur when applying terminology to experiences is provided in figure 1.2.1.

Figure 1.2.1 Shaw and Gaines, Model of Conceptual Structures (1995). Cited Berg (2002)

<table>
<thead>
<tr>
<th>Same Attributes</th>
<th>Terminology</th>
<th>Different Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same</td>
<td>Consensus</td>
<td>Correspondence</td>
</tr>
<tr>
<td></td>
<td>Experts use terminology and concepts in the same way</td>
<td>Experts use different terminology for the same concepts</td>
</tr>
<tr>
<td>Conflict</td>
<td>Experts use same terminology for different concepts</td>
<td>Contrast</td>
</tr>
<tr>
<td></td>
<td>Experts differ in terminology and concepts</td>
<td></td>
</tr>
</tbody>
</table>

According to Shaw and Gaines' model (1995), the transfer of information between people (defined as experts by the authors) is three times more likely to result in erroneous assumptions being made, than it

---

\(^{19}\) Rumsey was referring to terms relating to the 'immersive' qualities of a sound field as experienced by a listener.

\(^{20}\) Adjectives included; likeable, mean, good, narrow-minded and sincere
The graphical representation of listeners' auditory spatial experiences

is to result in consensus; whereby both experts are using the same terminology to represent a concept in the same way.

Hence subjective audio evaluation faces a challenge. It is tasked with obtaining from each listener a comprehensive description of their experiences as they appear when structured and represented within a context dependent language. This linguistic expression of a listener’s experiences can never be as complete as the experience itself. Moreover, the representation (when communicated) can be affected by (amongst other things) the knowledge and personal histories of the communicators, and the context of the communication. Communication - be it in written text, spoken word, or graphical depiction - is required at many stages within an evaluation, from the issuing of instruction to the collection of responses. At each of these stages, (from researcher-to-listener when presenting the subject to be studied; from listener-to-researcher when representing their experiences; and from researcher-to-reader when analysing what has been communicated and recording conclusions) the transfer of information between parties is susceptible to anomalies, each of which may affect the consequences of the research. However communication cannot be avoided since, through this mediated form, the researcher must attempt to understand each listener's unique experiences; the validity of the evaluation determined by the “adequacy of the researcher to understand and represent meanings.” (Banister et al., 1994, p143).

Historically, verbal language has been used in subjective audio evaluation to elicit from listeners representations of their experiences. The following section provides a brief overview of the verbal methods favoured in contemporary research and the information that may be obtained using these methods.
1.2.6 Eliciting verbal descriptions from participants in subjective evaluation

In 1985 Gabrielsson and Lindström conducted numerous investigations whereby they asked listeners to describe the perceived characteristics of different sound reproduction systems using adjective scales which had been developed in earlier experiments\(^{21}\). In these original investigations, Gabrielsson and his colleagues had asked listeners to rate (amongst other things) similarities between different sound reproduction systems and to describe these audio systems using their own verbal descriptions. Since, as Bech asserts, “only in a very few cases is it possible to state that the total auditory impression is determined by a single attribute” with most auditory impressions “based on a combination of a number of underlying auditory attributes” (Bech, 1999, p488), it was assumed by Gabrielsson et al. that the auditory experience of the listener would also comprise numerous dimensions. The authors therefore attempted to identify the constituent uni-dimensional perceptual attributes underlying listeners’ similarity judgements using multi-dimensional scaling (MDS). Results of the analysis established that a small set of contrasting experiences existed for listeners when presented with the different audio devices which remained constant throughout the investigations. These dimensions (and those found pertinent following the analysis of listeners own verbal descriptions) were given the following descriptive labels by Gabrielsson and Sjögren in 1979.\(^{22}\)

- brightness - darkness (brightness was opposed by dullness by 1985)
- clearness / distinctness (clarity was introduced besides clearness/distinctness by 1985)
- loudness
- sharpness / hardness - softness
- feeling of space (spaciousness was introduced besides feeling of space by 1985)
- fullness - thinness
- disturbing sounds (renamed absence of extraneous sounds before 1985)
- nearness

Through the specification of the various attributes, Gabrielsson and his colleagues had provided evidence of the multi-dimensional nature of reproduced sound, and how the individual dimensions within this multi-dimensional event could be verbally identified. To evaluate the appropriateness of each individual dimension (and their verbal labels) Gabrielsson and Lindström created 11 point (0 - 10) adjective scales for the attributes and asked listeners to use these when rating various loudspeakers (Gabrielsson and Lindström 1985). Results from their investigations indicated that although listeners tended to use different parts of the rating scales - this being partly attributed to differences in how the descriptive end points of the scales were comprehended - listeners were reliably able to discriminate between the loudspeakers.

Gabrielsson et al. had therefore identified verbal scales which could be used successfully to evaluate audio devices in terms of their timbral attributes or perceived distortion. However, although defining timbre in terms of several dimensions, the spatial qualities of the reproduced audio were only

\(^{21}\) See Gabrielsson et al. 1974 and Gabrielsson and Sjögren 1979

\(^{22}\) Although presented here in English, it should be noted that the original labels were in Swedish
considered with respect to one scale;23 spaciousness. Although this simplified view of spatial characteristics may merely have reflected the research's position in history, current research is positioned alongside the development of multichannel audio systems capable of delivering more obviously spatial information to the listener. Thus it is necessary for contemporary subjective audio evaluation to consider the listener's experience with respect to these spatial characteristics. But what is a spatial attribute?

The definition of a spatial attribute

In providing the following definitions of spatial qualities or attributes, an attempt is made to render these spatial characteristics distinct from other commonly elicited attributes of reproduced audio. However, more dedicated critiques of the terminology used for the various spatial attributes of reproduced sound have been provided by Berg (2002) and Rumsey (1998 and 2002).

Rumsey suggested that the term attribute be defined as "a characteristic quality of an object that one may use in describing it" (Rumsey, 1998, p123). More specifically, Bech described an auditory attribute as "a perceived characteristic of a sound stimulus, for example pitch and loudness" (Bech, 1999, p502). Thus, a plausible explication of a spatial auditory attribute would be a characteristic spatial quality of a sound stimulus. Accordingly, it becomes necessary to determine what constitutes a spatial quality, or spatial attribute. As recognised by Rumsey,

until recently there have been relatively few attempts in the world of reproduced sound to isolate any more detailed spatial attributes than all-encompassing ones, such as spaciousness, spatial impression, sound stage or stereophonic impression. (Rumsey, 2002, p652)

Spatial impression had earlier been defined by Rumsey as a multi-dimensional attribute, "having a number of [not yet precisely defined] sub-attributes such as 'envelopment', 'source width', 'image depth', 'stability', 'phasiness', and so on' (Rumsey, 1998). In 1993, Lehnert provided an overview of the terminology then believed to be allied to auditory spatial impression. Lehnert declared that "auditory spatial impression" consisted of both "auditory spaciousness" and reverberance. The author went on to characterise auditory spaciousness as the "spatial spreading" of an auditory event, typified by the apparent "enlarged extensions of the auditory image compared to that of the visual image" (Lehnert, 1993, p41). In 2001, Mason and Rumsey proposed spatial impression to be a multi-dimensional entity consisting of "location, dimensions, and other physical parameters of a sound source and the acoustic environment in which the source is located". (Mason and Rumsey, 2001, p1)

Letowski, used the term "spaciousness" to provide the macro level to his definition of spatial quality. Spaciousness was itself described as:

That attribute of auditory image in terms of which the listener judges the distribution of sound sources and the size of acoustical space. Spaciousness enables the listener to judge

23 'nearness' having been omitted in 1985
that two sounds which have, but do not have to have, the same pitch, loudness, duration, and timbre are arriving from different locations. (Letowski, 1989, p4)

For Berg, spatial qualities referred to “the three-dimensional nature of sound sources and their environments” (Berg, 2002, publication 2, p3). This could be more explicitly defined as comprising “those perceptual constructs that relate to the sensations of directionality, size, depth and width, of reproduced sources, groups of sources and acoustical environments” (Berg, 2002, p14).

Eliciting verbal spatial attribute descriptions from listeners

Not only did early adjective scales concentrate on timbral and distortion characteristics of reproduced audio rather than its spatial characteristics, but there was also a propensity for these scales to be provided by the researcher rather than elicited from the listener. Bech offered a definition of the differences between provided and elicited constructs in his paper from 1999:

Provided constructs are presented to the subjects and they are trained in using them in the given experimental context. Elicited constructs are generated by the subjects themselves without any influence of an experimenter. Results based on elicited constructs are therefore more... reliable as subjects are using their own words (Bech, 1999, p499)

For Gabrielsson and his colleagues, multi-dimensional scaling had uncovered the underlying perceptual attributes associated with the reproduced audio devices. However, the verbal terms used to define the eventual uni-dimensional adjective scales and to differentiate between these individual dimensions were primarily provided by the authors themselves. Regarding the reduction of a multi-dimensional entity into its uni-dimensional components Berg notes, “it is also important that the scales or dimensions defined are meaningful” (Berg, 2002, publication 1, page 3).

Since the experimenter is most likely to be the person able to define the factors of interest to them (ibid. p1), it is (arguably) valid for the experimenter to provide participants in a study with the attribute scales upon which they will be making their judgements. Adams-Webber’s 1970 paper on elicited and provided constructs in repertory grid technique provided a review of investigations in which participants had either used their own verbal descriptions of their own constructs or conversely, investigations where constructs had been provided for them by the investigator. Based on his literature review, the author concluded that individuals found their own systems of personal constructs more useful than provided constructs when structuring their own social environment. However, it was also noted that “normal subjects...exhibit approximately the same degree of differentiation in using carefully selected supplied lists of adjectives as when they employ their own elicited personal constructs” (Adams-Webber, 1970, p352). Landfield (1968) also looked at the use of provided constructs in the repertory grid technique to establish whether these were less meaningful to individuals that their own elicited constructs. Landfield based his conclusions on the hypothesis that listeners would demonstrate more extremity when rating using more meaningful construct scales. He concluded that listeners did indeed find their own constructs more meaningful than those provided by another. Landfield stated that his data gave “considerable support to the contention that one’s own language is
more acceptable to one than is the generalized language of others" (Landfield, 1968, p138). As noted earlier in this thesis (section 1.2.4), using provided scales may focus the attention of the listener towards attributes which do not correlate with their actual experiences when auditioning the reproduced audio. Consequently, results obtained using provided language may demonstrate poor validity for the listener. The following sections will therefore concentrate on methods which allow listeners to describe and structure their own experiences using their own terminology.

**RGT and the use of elicited terminology**

The Repertory Grid Technique (RGT) was introduced in section 1.2.4 of this thesis as a method of eliciting information from individuals regarding their own personal constructs\(^ {24}\). Briefly, the technique involves asking people to describe the ways in which two out of three presented events are alike and yet different from the third. Bech (1999, viewing the process from the perspective of an audio evaluator) declared that the technique thus enables the generation of constructs, or more properly, a number of terms representative of these constructs which can be used to fully describe the reproduced audio. For Berg, RGT "encourages personal reflection upon the qualities of the stimuli under examination, and definition of a personal set of constructs that differentiate between them" (Berg, 2002, Publication 1, p6). Consequently, RGT may be thought of as a method which enables an event to be understood in terms of how the individual experiences it.

Berg (2002) used RGT to involve listeners in the process of defining a set of verbal spatial attributes which could be used to structure their experiences of reproduced audio. Specifically the RGT method was employed to obtain from listeners descriptions of the similarities and differences between six different stimuli - each recorded with a variation in either microphone arrangement or multichannel reproduction mode. The characteristics elicited by Berg were believed to equate to a listener's constructs and, for Berg, "it is important to know what the constructs are, whether there is a common set, and also to adopt meaningful and appropriate methods of scaling that relate to the psychological continuum and to physical attributes of the sound field" (ibid. Publication 1, p1). Within the elicitation, Berg asked several participants to listen specifically to the three-dimensional nature of the sound sources and their environment, whilst other listeners were not restricted in this way. Even though unrestricted, this second group of listeners also provided verbal attributes which were predominantly spatial. Listeners rated the various reproduced stimuli on their own bipolar construct scales and Berg inspected the resultant information to detect similar constructs which existed for each listener and between listeners. Berg’s analysis led to the identification of the following groups of similar constructs for the listeners\(^ {25}\).

---

\(^ {24}\) Personal Constructs are the system of dichotomous contrasts used by the individual as they try to make sense of their experiences (see Landfield 1968 and Kelly 1963)

\(^ {25}\) Although presented in English, constructs were elicited and structured by Berg in Swedish
Group 1: *Authenticity or naturalness*, for example:

Group 2: *Lateral positioning and source size*, for example:
narrow sound source – wide sound source, a point – width, mono – stereo, limited – open,

Group 3: *Envelopment*, for example:
sound everywhere – sound from a part of the room, in the centre of the event – outside the event,

Group 4: *Depth*, for example:
frontal depth – rear depth, sound source in the loudspeaker – sound source between the speaker and me, sound source is placed on a line – more depth

In a later analysis (see Berg, 2002, publication 2) listeners' original constructs were further categorised as: localisation, left-right and front-back; depth / distance; envelopment; width; room perception; externalisation; phase; source width; source depth; detection of background noise; and frequency spectrum. Berg noted, however, that his analysis showed no consistent division of the attributes into specific groups, with several attributes found in more than one group. Berg suggested this could be a result of the listeners using the same terminology for different attributes, or an indicator that the stimuli were too complex; exciting many dimensions simultaneously. Berg further highlighted the possibilities of bias when attempting to derive a common set of spatial attributes from individually elicited constructs acknowledging that “the relatively free, and thereby low-bias, approach at the elicitation stage in this experiment results in more dispersed verbalisations at the stage of analysis” (2002, publication 2, p6). But, interpretation has to be done by someone and, as Berg declared, “an advantage with this [method of elicitation] is the availability of relatively unbiased original data, for the event that other methods of analysis will be used later on” (ibid.). Thus Berg believes his work to be “a valid starting point for designing new experiments aimed to investigate the aspects of spatial sound reproduction.” (ibid.)

**The use of Descriptive Analysis in spatial audio evaluation**

In 1999, Bech proposed that Descriptive Analysis (DA) could be used as a beneficial tool for establishing the multi-dimensional spatial characteristics of reproduced audio devices as determined by a panel of listeners. Already established as a method of evaluation in food and consumer sciences (see Meilgaard et al. 1999 and Hootman, 1992) DA required little additional alteration before being applied to the analysis of reproduced audio. For Bech, the purpose of auditory Descriptive Analysis was to:

- identify individual auditory attributes
- devise methods for obtaining a measure of the magnitude of sensation for each attribute
- establish the relation between the auditory attributes and the total auditory impression

The descriptive analysis process was outlined by Bech as consisting of the:

- selection of subjects
The graphical representation of listeners' auditory spatial experiences

Chapter 1 - Introduction

- development of a descriptive language (free from leader influence), and training of subjects
- quantitative reporting of evaluations, analysis and presentation of the data

When selecting listeners, Bech ascertained that, as well as basic listening aptitude, listeners should be able to quickly recall verbal attributes and apply these to different products. With respect to these requirements, competence could be assessed using pre-screening questionnaires, acuity tests and (more specifically to DA) an interview to get a feel for the listener's communication and other social skills.

The development of the descriptive language in DA is itself a complex process involving several stages. The first step is for the panel of selected listeners to generate (with minimum intrusion from the researcher) a list of terms which may be used to describe the characteristics of a representative sample of audio stimuli. Panelists then remove duplicate words from the original list of terms and group the remaining terms according to similarities in the characteristics being described. The aim throughout the language development process (during which time the panel's language continually evolves) is to create a comprehensive set of verbal terms, capable of describing all the characteristics of the representative audio samples using the minimum number of words. Within themselves, these verbal descriptors need to: (i) discriminate between stimuli and relate to the physical measures defining the stimuli (ii) be orthogonal and use singular rather than holistic terminology (iii) relate to concepts that influence consumer preference decisions (iv) be precise and reliable (v) generate consensus among (and be unambiguous to) the panel (vi) have communication value and not use jargon, and finally (vii) the verbal descriptors should relate to reality. After developing their descriptor set, listeners are required to establish descriptive end points for attribute scales using their verbal language. These scales can then be used in the assessment of reproduced audio, specifically in defining the extent with which each reproduced event correlates with each verbal attribute.

Training panelists to use their verbal descriptor scales also forms an integral part of the language development phase. Panelists are initially introduced to simple audio examples which represent the developed verbal terminology and demonstrate a wide spread in intensity on the different scales. Following this initial training, panelists begin to rate less obviously different examples. According to Bech, results of these trials may be used to verify the response system and to examine the consistency of both panel and individual listeners. Following training and validation, the resultant scales and trained panel can commence the evaluation of audio devices.

Because of the level of discussion involved in the generation of the verbal terms and evaluative adjective scales, Bech noted that descriptive analysis may suffer from the adverse influence of the researcher (panel leader), or bias resulting from the in-depth discussion of the panelists. Similarly, Berg described the DA method as "something of a cross between provided and elicited constructs, as subjects are influenced and perhaps biased by each other", yet it is still capable of providing the listeners with an "opportunity to influence the choice of scales and their definitions" (Berg, 2002, publication 1, p2).
Zacharov and Koivuniemi (2001a, 2001b) and Koivuniemi and Zacharov (2001) devised ‘ADAM’ (Audio Descriptive Analysis and Mapping) to obtain an "understanding of the multidimensional structure of the perceptual space, in the form of salient perceptual attributes" (Zacharov and Koivuniemi, 2001a, p273). As in standard descriptive analysis, ADAM involves the complex process of developing a descriptive language and set of magnitude scales which can be employed in the attribute rating of reproduced audio events. In a specific investigation using ADAM, 12 panelists generated an initial list of 1400 attributes to describe audio stimuli representative of a wide range of acoustic environments. These original attributes were reduced to a more manageable 532 terms, a process which was possible since words could be omitted which had the same stem. Group discussions took place over five weeks to give panelists time to argue or concur about the descriptors which should be included in the common descriptive language. The outcome of the language development phase was a set of verbal attributes which, although requiring further analysis and refinement, was believed by Zacharov and Koivuniemi to provide valuable information about the salient percepts of spatial sound reproduction. The terms included the following spatial attributes:

- Sense of direction (how easily the location of events can be discriminated)
- Sense of depth (how strongly the sensation of distance is perceived)
- Sense of space (how well the space where the recording was made is perceived)
- Sense of movement (whether a sound source is perceived to actually move in the sound space)
- Penetration (a positive value means that spatial information seems artificial; the sound seems to originate very close to, or even inside, the listener’s head)
- Distance to events (the actual distance to where the sound events appear to originate)
- Broadness (how wide an area the perceived sound event seems to have)
- Naturalness (how well the perceived events conform to the listeners’ model of realism)

Concluding remarks

Contemporary subjective evaluation has increasingly looked at verbal elicitation methods to provide information about the spatial attributes of reproduced audio as experienced by the listener. These methods have the advantage of increased validity over more traditional methods (when examined from the perspective of the listener), since it is the listener’s own words, or those of a panel of listeners, which are present in the resultant descriptive terminology or attribute rating scales.

No matter how useful verbal information is, it is worth remembering that verbal communication is not our only means of exchanging knowledge, and this is especially true when what is being communicated is not inherently verbal - as is the case with listeners’ experiences. It may therefore be possible to obtain additional information about the spatial characteristics of reproduced audio if an additional communication medium is employed. As Levy declares,

---

26 see Zacharov and Koivuniemi (2001a, pp278-282) for a detailed description of the stimulus set.
27 Terminology was elicited in Finnish, the native language of the participants.
28 A process which is thoroughly documented in Koivuniemi and Zacharov (2001).
language makes one sensitive to certain stimuli and not to others, to the ways in which language seems to mould our ways of thinking and dealing with events. This it would seem, could account for the fact that when different languages are invoked to deal with a problem, we find different items of information brought into the foreground (Levy, 1963, p19).

Novitz concurs with Levy, affirming that the way in which events are described influences not only a person's attitude to these events, but also what they notice about them; the particular features which are singled out for attention (Novitz, 1977, p109). He continues, "One's ability to describe an object, and more particularly the way in which one describes it, often affects one's ability to recall it" (ibid. p119). The malleable nature of an event when described in different media is further discussed by Levy, who asserts that "while some events seem to resist description in more than one language system, others are not so recalcitrant. In each case however, we select the language which we expect will maximise our ability to deal with the problem at hand" (Levy, 1963, p12).

As previously noted, linguistic descriptors can be a limiting factor in an individual's ability to express their experiences. This situation occurs in part because an individual's experience of their environment is continuous (although selective), but the linguistic descriptors available to the listener when describing their experiences are somewhat limiting. If we are to place our confidence in Levy's assertion that "events may be viewed from many perspectives and each may suggest a different interpretation" (Levy, 1963, p92), one assumption is that using an alternative medium, when representing a listener's experiences of their environment, enables the researcher to elicit alternative attributes of this experience to those obtained through verbal communication. As Levy contends, "what we are in essence trying to do is take the individual event along various different paths until all interpretations of it have been exhausted" (Levy, 1963, p92).

Thus, by altering the communication medium, the message being communicated may itself be altered, providing the researcher with an opportunity for acquiring different and novel knowledge about the experiences of an individual. The following section provides examples of instances where alternative communication media have been employed in research on spatial perception. In doing so, the section also considers the necessary requirements for communicating experiences in a different language.
1.2.7 Alternative methods for representing spatial percepts

A different approach calls for a different lexicon (Kelly, 1963, xii)

The egocentric nature of directional spatial experiences

Hart and Moore (1973) asserted that when an individual endeavours to represent externally their internalised perception of a physical environment, they employ a reference system which allows for their spatial orientation within this environment. Similarly, Wickens and Prevett explained in their paper (Exploring the Dimensions of Egocentricity in Aircraft Navigation Displays, 1995) that any navigator through any space, whether real or virtual is required to perform two generic tasks: (i) local guidance, which involves maintaining the desired path and (ii) global awareness, which requires a knowledge of where things are in their surrounding environment. When identifying how an individual relates to objects and spaces around them, Hart and Moore believed that individuals used an egocentric system of reference. Howard and Templeton (1966), defined egocentric as the positioning of an object with respect to the body of the individual with no reference being made to any external point.

Since our spatial awareness is considered egocentric, Evans suggests that participants in listening evaluations (where judgements about directional experiences are sought) should be allowed to respond with reference to themselves, stating

if we are giving names to particular directions, do not assign names based on (to the listener) abstracted angles of azimuth and elevation. Instead, name the directions according to some pattern that listeners will be familiar and comfortable with. In general, this system of response should be inherently egocentric; effectively taking its reference from the position that the listener perceives him or herself to be. (Evans, 1998, p3)

It is useful therefore, to start with egocentric response methods when exploring alternative media for representing an individual’s experience of directional spatial information.

Eliciting directional information from participants using non-verbal methods

According to Montello et al. (1999), there are a wide range of egocentric response techniques to choose from when studying directional knowledge. Freeman (cited in Hart and Moore, 1973) states that direction is “represented in the mind” in terms of moving the body, either through turning the head or pointing, both methods aligning the individual with the required direction. It was similarly suggested by Montello et al. that participants in directional estimation investigations can point (either with their hands or some other object), turn their heads or eyes towards the current direction, rotate their bodies or, in some cases, walk along a particular course. An investigation by Montello et al. compared two different directional estimation methods, one of which was egocentric, the other external to the participant. The study involved a group of 24 participants who were either blindfolded or had their sight partially restricted for its duration. The participants were instructed to indicate the location of a particular visual object (the position of which they had previously memorised) within a 360° field of view.
The graphical representation of listeners' auditory spatial experiences

Chapter 1 - Introduction

reference using either an external 'manual dial' device or by rotating their body towards the direction of the item. Unusually, the results indicated that eliciting directional information using a manual dial is as effective as the more egocentric body-rotating method. Montello et al. contended that this result was adversely affected by the performance of the blindfolded participants when rotating to face the estimated location of the object. The authors suggested that participants can estimate direction quite well when they are able to see their feet and the surrounding floor, but once blindfolded have to rely on their short-term vestibular memory to orientate themselves: If a person falters whilst rotating, they will have no access to their initial heading and thus their position with respect to objects becomes confused. The equivalent results for the manual dial technique showed no such errors, indicating that participants can maintain orientation at all times during pointing. However it should be noted that participants in this part of the study may have relied as much on their memory of object placement as their ability to indicate location using the different estimation methods.

The results of the Montello et al. study contradict the findings of a similar study by Haber et al. (1993). Haber et al. asked 20 blind participants to estimate the direction of a pure tone audio signal reproduced over one of five loudspeakers within a 180° field of reference. Nine directional elicitation methods were employed by the authors, including three egocentric pointing techniques, two less egocentric methods (where the participant pointed with an external device) and four further methods requiring the participant to use an entirely external estimation method. Two of these latter methods involved a similar manual dial to that employed by Montello et al., which was either fixed at the participant's waist, or positioned on a table at waist height. A further method used the same manual dial board, but with the pointer replaced by a piece of paper so that the participant could draw a line to indicate the perceived direction of the sound. Finally a single verbal estimation method was used, whereby the participant was asked to indicate the location of the object as if it were a position on a recumbent clock face with 12:00 occupying the location straight ahead. For Haber et al., the greatest accuracy was obtained using the egocentric and less egocentric pointing methods, with their two manual dial, drawing and verbalisation techniques eliciting more variable results.

Sources of ambiguity in non-verbal response methods

In his review of techniques for obtaining responses in directional listening tests, Evans presented several sources of ambiguity which had been identified in previous research. The first method to be critiqued by Evans was elicitation using azimuth and elevation notation. Evans noted that asking listeners to describe sound by verbalising degrees of azimuth and elevation could produce inaccurate results, since listeners would not necessarily be familiar with this type of verbal terminology. However, even with training, listeners could still have difficulty in translating their "directional percept into a correctly formatted angular response" (Evans, 1998, p2). Thus Evans believed the "cognitive overhead associated with giving non-intuitive responses might yield an increased rate of general mistakes in communicating angular estimates" (ibid. p3). To provide an alternative to verbalisation using angular

29 The manual dial device consisted of a smooth circular piece of cardboard with a rotatable length of wire attached to indicate the estimated direction of the object.
In Evans’ study, listeners had been asked to estimate a sound’s direction with respect to hourly positions on the clock-face. Hence the listeners were required to estimate the position of a sound to within the nearest 30°, somewhat limiting listeners' responses. Although Haber et al. encouraged participants in their study to use a continuous scale when indicating the estimated direction of a sound on the clock-face, the units of measurement most frequently used were those of 15 or 30 minutes. This finding would suggest that the participants were unable to use the continuous scale or were unable to translate the perceived direction satisfactorily into the verbal position. Alternatively this finding could indicate a trend in individuals for time-telling in a certain manner. Research by Nielsen indicated that participants like to quantise their responses into “nice” values. Nielsen asked his participants to estimate the distance to a particular sound (reproduced over loudspeakers) and to indicate the placement of this sound using a computerised graphical response form. Results of his experiment led Nielsen to believe that “when subjects hear a sound they tend to place the marker on whole numbers or simple fractions thereof” (Nielsen, 1991, p9).

Even egocentric elicitation methods are not without problems. According to Howard and Templeton (1966), in a pointing experiment, accuracy is not only a factor of how well the individual can localise the sound, but is limited by the ability of the participant to position their hand with relation to their experience. This finding is in agreement with the Haber et al. study where a source of constant error was in the positioning of the pointer, with participants finding it difficult to maintain a straight arm. This error is accentuated when an external pointing device is used, as the participant has to control both their own movement and that of the device. Furthermore, as recognised by Montello et al. “material objects and landmarks have width; they subtend several degrees of azimuth in some cases” (Montello et al. 1999, p990), accordingly this factor could contribute ambiguities to all directional estimates including egocentric pointing methods.

36 It could be argued that a participant’s natural ability will be a limiting factor in their use of any response method (be it verbal, graphical or egocentric).
Although the reviewed literature is by no means conclusive, it appears (for the most part) that participants estimate the perceived location of an unseen object best when using egocentric or intuitive pointing responses. Yet, as indicated by Haber et al., pointing has its limitations; "pointing alone typically defines only direction" (Haber et al. 1993, p36). Pointing is therefore not an immediate choice of elicitation method when what is desired is an adequate description of the total auditory space as experienced - it appears immediately unintuitive to point to an auditory object's size for instance. In such cases, an alternative method of elicitation must be employed. According to Evans, "an extremely elegant mechanism for listeners to give their responses with complete directional freedom is to use a graphical representation of apparent direction" (Evans, 1998, p7). Furthermore a pictorial description enables a two-dimensional representation of the auditory scene to be displayed in its entirety.

It is interesting to consider whether a pictorial, graphical language could be used to elicit from listeners, representations of their experiences of spatial attributes in reproduced sound.

Translating egocentric spatial experiences into graphical representations

When identifying the viability of a graphical medium for communicating a listener's perception of spatial audio, it is necessary to consider whether the listener will be able to map an event which occurs in auditory space onto a space defined in the visual modality. In brief response to this consideration, Auerbach and Sperling (1974) provided a review of the literature on directional information and conducted an investigation which confirmed the existence of a 'common space' dimension. 'Common space' was defined by the authors as an underlying single dimension existing independently of the modality (either visual or auditory) of the stimulus having the direction. Thus, rather than existing on two separate direction dimensions, the authors concluded that auditory and visual directional information was immediately referred to a common spatial framework. The existence of the 'common space' led Auerbach and Sperling to assert that no translation would be required when comparing visual and auditory directional information.

Woszczyk likewise noted how auditory and visual perception complement each other. In his paper on the Quality Assessment of Multichannel Sound Recording the author suggests that a "sonic picture" is constructed by a listener to provide a model of the world. While they are listening "the listener creates in his mind a visual representation of auditory scene which consists of sound sources and of acoustical environments, including their complex spatial and qualitative relationships" (Woszczyk, 1993, p198). Woszczyk also mentioned a previous informal investigation conducted in two concert halls to assess the importance of a vertical dimension in music perception. In this investigation, orchestral concerts were attended and visually described. Woszczyk noted that, although similar images of instruments were heard in the different concert halls, these images varied substantially from instrument to instrument. He writes:

The total image produced by the orchestra was like an ensemble of pulsating shapes harmonized with the music. Some images were tall and narrow (like flutes and woodwinds)...or sharply defined as point sources (like trumpets), while some changed...
with dynamics and intensity (like strings which produced tall and narrow image at pianissimo, but low and wide image at fortissimo), still others occupied the entire vertical and horizontal space. (Woszczyk, 1993, p200)

Because of the capacity for individuals to visually represent their auditory percepts, Woszczyk contended that a graphical tool may be used to record the characteristics of sonic pictures, and - in this way - help in the assessment of multichannel audio by identifying, amongst other things, balance problems, lack of symmetry, poor separation or incomplete surround information. Woszczyk believed that only basic drawing skills would be needed for the task of drawing the perceived auditory image. He further suggested the use of a two-dimensional response form.

Accordingly, not only does using a graphical response medium require the individual to switch modalities and move from an egocentric reference to a graphical plan on paper or computer screen, but a further consideration is the translation of three-dimensional space into two dimensions. For Evans, although listeners can be given a diagrammatic representation of the listening environment, (believed by Evans to be a generally straightforward means of describing the relative directions of sound sources) this representation can still only be a two-dimensional projection of three-dimensional space. Evans continues, “for the listener, translating the perceived direction into such a representation may encounter similar problems as expressing a direction in terms of azimuth and elevation angles” (Evans, 1998, p7).

Difficulties of translating three spatial dimensions into a two-dimensional representation were also acknowledged by Amheim who believed that, no matter how useful plan drawings could be for inferring information about internal representations, they should not be seen as a simple translation from perception, as “misinterpretations are inevitable if the picture is considered a more or less correct replica or derivative rather than a structural equivalent of the object in terms of the medium” (Amheim, 1956). Amheim suggested that when three-dimensional space was represented in two dimensions, the simplest and most characteristic aspects would be depicted. This was exemplified in Amheim’s study of children’s scribbles; where space existed only in two dimensions, nothing distinguished between a flat or a voluminous object and the spatial qualities of a dinner plate were treated no differently from those of a football.

The problem of visually depicting internal spatial perceptions was also studied by Shemyakin (cited in Hart and Moore, 1973). The author established that when young children draw a plan of some locality familiar to them, they usually do so by means of a route map. These routes typically begin at the edge of the paper closest to their body and are drawn away from the child so that left and right turns occur as they would be experienced. When Shemyakin asked the children to draw routes towards themselves - essentially inverting the plan - it was noticed that the number of errors increased.

Describing the spatial attributes of a sound event does not explicitly suggest the drawing of route maps, however, the listener may be asked to estimate and graphically represent the perceived location of an auditory event within a space. Since the placement of this event can be considered a form of mapping,
it is prudent to contemplate the implications of using a two-dimensional plan. Further to the direction in which the plan is presented to the listener, participants may also find it easier to construct a representation of what they have heard when they have reference to fixed landmarks - for instance the position of the walls within the room or the position of their outstretched arms. Hart and Moore (1973) attributed this requirement to each individual’s fixed system of reference, whereby orientation within space is partially co-ordinated by the use of landmarks.

Using a graphical medium in subjective audio evaluation

Previously, graphical methods have been used for the provision of information regarding the perceived location of a sound within a space. Researchers who have used a graphical method in this manner include Nielsen (1991 and briefly outlined above) and Møller et al. (1996) - who used an approximation of the method proposed by Nielsen, by asking listeners to indicate (with a digital pen on a computerised pad) which of 19 visible loudspeakers was perceived to have produced a sound. A graphical response method was also employed by Wenzel (1999, for estimating the azimuth and elevation of, and distance to, a virtual source of broadband noise) and Martin et al. (1999) who, in their study of phantom image localisation in multichannel reproduction systems, instructed listeners to indicate the perceived location of each phantom image by placing a dot on a computerised map of the listening room.

Moving away from purely directional information, Blauert and Lindemann (1986) asked the 12 listeners in their study to map - by drawing on a response form depicting a representation of a head viewed from the top and the front - the single or multiple auditory events they perceived during the presentation of various noise signals. Listeners were required to draw the edges of the stimuli (as perceived) on their paper response forms. The resulting circles on each response sheet were filled with black, placed one by one in front of a camera and exposed onto the same piece of film. The result was a density plot showing darker regions where more listeners had depicted the stimuli and lighter regions where fewer listeners had represented the sounds. Martens (1999) used a graphical method to obtain informal responses from participants in an experiment looking at the influence of low frequencies on a listener’s spatial perception; more specifically the nature of the perceived location, shape and “spatial extent” of a reproduced sound when two subwoofers were employed rather than one. The response form provided by Martens consisted of a representation of the speakers, a line between these speakers and a line from the listening position to the speakers. By implying the range of acceptable drawing with this blank form, and issuing instructions about what to draw, Martens noted the experimenter bias which could influence listeners’ responses; a necessary consideration for all subjective audio evaluations regardless of the language employed.

Recently, Usher and Wosczczyk (2003, 2004) investigated the use of a graphical response tool for visualising and revealing perceptual spatial differences in reproduced sound scenes. In brief, Usher and Wosczczyk’s graphical user interface (GUI) enables the evaluation of multichannel loudspeaker systems by mapping the spatial characteristics of sound regions onto a two-dimensional plan (viewed from above) using ellipses. In their earlier investigation (2003), listeners were asked to describe “hot-spots”
The graphical representation of listeners' auditory spatial experiences

Chapter 1 - Introduction

(or "local regions of certainty about the listeners perception of a sound image at a particular location", ibid. p157) within sound regions. Results of this investigation satisfied the authors that the graphical tool could be used to represent source-related (as opposed to environment-related) sound images. In 2004, the GUI was employed to investigate perceived spatial extent and "definition" (whether a reproduced sound is defined, diffuse or fuzzy). The authors concluded that the tool could be used to describe a listener's perception of both attributes.

Thus, graphical response media can be - and have previously been - used to elicit representations of spatial attributes from listeners in subjective audio evaluation. It has been identified that the medium may be used to provide directional information about a reproduced audio event, the spatial extent and shape of a sound and (more recently) regions of perceptual certainty and definition. Additionally, ways of minimising errors due to the necessary translations from egocentric reference to external response and three-dimensional auditory environment to two-dimensional visual plan have been identified. Methods include adjusting the view of the response sheet to reflect the listener's perspective and the addition of landmarks on the response sheet to provide scaling information.

It should, however, be recognised that in earlier studies there has been little evidence of the systematic evaluation of the graphical response method itself. Rather, researchers have tended to accept the medium as suitable or have included graphical responses as an informal approach to viewing listeners' perception of reproduced audio. Typically the medium has been used to obtain from listeners specific information about auditory spatial attributes (for example, direction or spatial extent). There is a notable distinction that needs to be made between obtaining responses using a medium and eliciting from listeners a representation of their experiences using a graphical language: The former obtains information about a specified reference object, whilst the latter explores and identifies the experiences which can be structured and described using a graphical medium. It should also be noted that listeners have frequently been provided with the graphical descriptive terminology to use when describing the spatial attributes under investigation, rather than being allowed a more free response style.

It has already been mentioned that enabling listeners to use their own language when structuring and representing their own experiences allows for a greater understanding of these experiences. Verbal language has already been used to elicit from listeners information about their underlying experiences, but altering the medium from verbal to graphical provides researchers with the opportunity for uncovering novel knowledge from a different linguistic perspective. And, as Novitz maintains,

pictures play a special and very important role in communication. We know, for instance, that by using a picture one can reveal in a matter of seconds what it would take minutes to describe. A picture appropriately placed can often disclose in a flash what many pages of writing would fail to convey. And this, or course, suggests that pictorial describing, reporting or explaining differs radically from its verbal counterpart, since in some cases at least, it is much more effective than using words. (Novitz, 1977, p68)
1.3 Summary

In subjective audio evaluation, meaning has historically been associated with objectivity or reliability. Consequently, the rigorous control of variables (other than those under investigation) has typically been advocated to obtain from the listener a meaningful (objective) measure of the perceived qualities of the object under study. Included amongst these extraneous factors are environmental variables - such as the loudspeaker location, listening position and programme material (stimuli) - the investigation setting (for example, order effects, visual bias and experimenter bias) and the listeners themselves. To obtain meaningful information, listeners have been trained to improve the correctness of their responses (how well these correlate with known physical parameters of the object) and have been selected for their listening expertise. Thus, for the majority of researchers working in the field of spatial audio - and specifically the subjective evaluation of spatial audio - it is the external audio device that provides the motivation for research. Questions are built around this external focus, listeners are asked for their perception of the object, and responses are read or viewed with respect to furthering the experimenter’s knowledge of the audio device.

Rather than associating meaning with a reliable measure of the object being studied, an alternative perspective is to position meaning with the subject of the subjective evaluation. Meaning here is constructed by the individual when presented with the external stimulus. Thus, instead of asking the listener to characterise the external object as known by the researcher, the “burden of discovery” (Kelly, 1963) is placed on the researcher. For a subjective evaluation to be considered meaningful in these circumstances, the researcher is encouraged to acknowledge the validity of the listeners’ individual experiences and to explore and understand these in more detail. Here, accuracy is assessed by the researcher’s ability to understand the listeners’ experiences. Within this more qualitative research ethos, the rigorous control of the listener is less effective, since the listener - as originator of meaning - should be permitted the freedom to respond to their valid experiences as experienced. These experiences may remain hidden if not chosen as the focus of the investigation by the researcher, or may be trained out, or omitted through the process of listener selection. Hence, if the listener is too highly controlled, responses may only reflect the researcher’s prophecy of what should happen in the investigation and consequently (although considered meaningful for the researcher) these responses can have poor validity for the listener.

This is not to say that qualitative research is always meaningful. Controlling extraneous variables of the environment and investigation setting can be just as beneficial in qualitative subjective evaluation as in more positivist research. Furthermore, qualitative research is ill-advised if (amongst other considerations) it fails to: (i) adequately investigate unusual responses; (ii) understand limitations concerning the population being studied; (iii) acknowledge the researcher’s role in data acquisition (and resultant bias in the interpretation of these data); (iv) consider returning results to respondents for feedback.
Evaluating the accuracy of a subjective audio evaluation is therefore not straightforward, since the criteria for the evaluation - either exploratory (subject-focused) or confirmatory (object-focused) - will determine where the priorities for the evaluation should lie and which controls should be in place. Should the listener be required to provide information about an audio device as known by the researcher (confirmatory), it is prudent to control all variable aspects of the investigation (including the listener themselves) to ensure the reliability or the objectivity of the results. However, if the researcher is attempting to explore and understand the listeners' experiences in greater detail, less restrictions should be placed on the listener and the onus for providing an adequate description of the listeners' (originally valid) experiences should lie with the researcher.

However, acquiring knowledge about listeners' experiences is not a simple process, since experiences are distinct from the languages which enable their structuring, differentiation, description and communication. Effectively, what is communicated in language is not a listener's experience, only a representation of this experiences in a communication medium. Here, subjective audio evaluation faces a challenge. Whatever the medium, the application of linguistic sign to experience is arbitrary (only achieving significance via cultural consensus) and the use of these signs within the creative context of communication is itself variable. Furthermore, as previously noted, the linguistic representation of listeners' experiences can never be as complete as the experiences themselves. Thus there may be difficulties finding the right words to express experiences (language unable to represent the subtleties of thought), or the chosen language may mould experiences or prevent listeners from describing difficult percepts. Moreover, it is difficult to state with any certainty whether the participants in an evaluation will use the same language in the same way as one another or the researcher, since language has a personal meaning for each communicator.

Accordingly, from the listeners' individual understanding of what is required of them, through the representation and subsequent communication of their experiences, to the researcher's comprehension of what has been communicated (and the description of their understanding using their own personal constructs) it is clear that what is later digested by the reader - who may, or may not, fully understand the language of the communication - cannot be thought of as a complete description. Rather, what is attempted is the effective description of the listeners' experiences using a descriptive language, a representation which makes an understanding of the listeners' experiences possible when communicated.

The researcher has therefore to provide listeners with a means of describing their experiences. As a result, it becomes necessary to raise further questions, such as what are the experiences of reproduced audio that need to be communicated and how are these experiences best represented?

The development of multichannel audio has brought with it an opportunity for listeners to experience a greater diversity of spatial information than previously attainable. Thus the focus for much contemporary research has moved away from identifying and evaluating timbral characteristics and
moved towards distinguishing the underlying percepts that constitute the spatial attributes of reproduced audio. Moreover, research has undertaken to label these perceptual dimensions using either provided or elicited verbal terminology. Verbal language provides listeners with a means of structuring and communicating their experiences of spatial audio. Furthermore, when elicited, this language supplies the researcher with information about salient perceptual spatial qualities, information which can later be used to evaluate the spatial characteristics of different audio devices. But, verbal communication is not the only medium open to the listener and researcher in subjective audio evaluation. Correspondingly, it may be possible to obtain novel knowledge about listeners' experiences if these experiences are represented using an alternative medium.

Graphical communication media have been used previously as an alternative to egocentric pointing methods when estimating the direction of a reproduced audio stimulus. Although requiring the listener to translate between an egocentric reference and an external perspective, these pictorial methods enable a two-dimensional representation of a three-dimensional auditory environment to be displayed in its entirety (within the constraints of a response form). Thus the medium can also provide information about the spatial extent and shape of a reproduced sound. However, even though graphical responses have been solicited in subjective audio evaluation, studies have shown little evidence of the systematic evaluation of the graphical medium itself; with the research typically conforming to a confirmatory rather than exploratory model. In other words, such research has focused on using a graphical medium to provide information about specified spatial attributes (defined verbally), rather than the acquisition of knowledge about the listeners’ experiences using a different communication medium. Moreover, the style of graphical response has - more often than not - been prescribed by the researcher, rather than elicited from the individual listener or developed by a panel of listeners. Essentially, rather than exploring how a listener can structure their experiences using a graphical language, researchers have asked listeners to use the medium when representing specified attributes of an external reference. It is a subtle, yet nevertheless important distinction; a distinction that needs to be made, and one which provides the foundations for the remainder of the thesis.
1.4 Statement of research objectives

The research reported in this thesis seeks to explore listeners' spatial experiences of reproduced audio in order that these auditory experiences may be understood by the researcher. It is envisaged that such an exploration will provide the means of discovering valid percepts for the future evaluation of spatial audio quality and the further development of multichannel audio devices. However, the evaluation of audio devices will not form the primary aim of this work. Rather, since listeners' auditory spatial experiences need to be structured, represented and communicated to the researcher using a communication medium, the thesis considers the development of a graphical language for fulfilling this purpose.

It is the aim of the research to meet the challenge of developing a descriptive graphical language which enables listeners to structure and represent their auditory spatial experiences of reproduced audio. However, it is not only necessary for the listener to be able to describe their experiences using a graphical medium; if the researcher is to obtain an understanding of the listeners' experiences, these need to be communicated to the researcher via the listeners' graphical representations. Moreover, to be useful to the wider research community, the researcher's understanding of the listeners' experiences needs to be further communicable. Accordingly, the specific objective of the research can be elaborated as follows:

*To develop a descriptive graphical language which enables the structuring and representation of listeners' auditory spatial experiences in order that these experiences may be understood when communicated to the researcher.*

Achieving the research objective involves a descriptive process whereby the aim is to arrive at a position where an understanding of listeners' auditory spatial experiences is possible. The process necessarily commences with the listener, for it is the listeners' experiences that require exploration. Thus, for this thesis the listener is the originator of meaning. Since a listener's experiences cannot be directly accessed by either researcher or listener, the next step in the descriptive process is for the listener to structure and represent their experiences within a communication medium. The choice of medium moderates the message being communicated, and the decision to use a descriptive graphical language enables an alternative perspective on listeners' experiences to be obtained. Correspondingly, the next stage of the descriptive process involves the communication of the listeners' experiences to the researcher using the graphical language.

The successful completion of the descriptive process occurs when the researcher can comprehend the listeners' auditory spatial experiences via their graphical representations. Having acquired an understanding of these experiences, this knowledge can be used to further communicate listeners' experiences to the wider research community and enables any use of the language in the subsequent evaluation of audio devices to be based on valid foundations. This final stage is beyond the scope of the current thesis, which goes as far as to evaluate the developing graphical language.
Figure 1.4.1 displays the descriptive process as a graphical model. As visualised in this model, an audio stimulus (St) can give rise to any number of different experiences in a listener (Exp¹ - Expⁿ) and these unique experiences (rather than the external stimulus) provide the primary topic of investigation for the research. When viewed with respect to the model, the research objective can be separated to form two integral parts within the descriptive process. The first of these is the structuring and representation (R) of listeners' auditory spatial experiences (Exp¹) within the graphical language. The graphical representation (R) is then used to communicate the listeners' experiences (Exp¹), and to enable the researcher's understanding (or comprehension, 'C') of these experiences. A key to the descriptive process model is provided in appendix 1.

To successfully fulfil the research objective, the researcher needs to acquire an understanding of listeners' auditory spatial experiences. To meet this objective, the successful completion of both stages of the descriptive process is imperative, since a failing at either stage would disrupt the progression between the listeners' experiences and researcher's comprehension. Specifically, it is necessary for the graphical medium to enable the structuring and representation of listeners' auditory spatial experiences of reproduced audio - progressing from (Exp¹) to (R) in the descriptive process - and subsequently to allow for the communication of listeners' auditory spatial experiences to the researcher, progressing from (R) to (C) in the model.

It is therefore necessary to establish a measure of success to identify if the research objective has been attained. But how can this success be measured? In essence, at what point can a graphical language be said to have successfully structured, represented and communicated the listeners' auditory spatial experiences of reproduced audio, and when can the researcher's understanding of a listener's experiences be identified as adequate?
1.5 Evaluating the effectiveness of the descriptive graphical language

In subjective audio evaluation, quasi-objective experiments are common. Here, the listener provides a measure of the stimulus under study and the effectiveness of the measure is verified by identifying whether listeners can consistently describe statistically significant differences between stimuli. Since, in this thesis, meaning originates with the listener (and the listeners' auditory spatial experiences are accepted as valid), it is not sufficient to correlate the listeners' graphical descriptions with the external stimulus as known by the researcher, as the listeners' experiences of stimuli may differ from those of the researcher.

Consequently, it follows that the effectiveness of the descriptive language should not be exclusively evaluated by correlating listeners' responses with physical characteristics of the stimuli as known by the researcher. This confirmatory style of analysis can ascertain where listeners are consistent in their graphical depiction when repeatedly presented with the same stimulus; or when there are differences in representation which correspond to physical differences in stimuli. But, quantitative methods are less appropriate for determining the effectiveness of the descriptive language where there are apparent inconsistencies in listeners' representations, or when differences in description do not relate to the physical stimuli.

In these anomalous situations, where quantitative methods would suggest the inadequacy of the listener, qualitative methods provide a suitable means of evaluating the effectiveness of the descriptive language. Miles and Huberman (1994) proposed that a check on the meaning of any outliers (exceptions) within a data set would help build a better explanation of the data. Furthermore, the investigation of any surprising results was also undertaken. As Miles and Huberman explained, "when you are surprised, it usually means that something has occurred well outside of your expectations" (ibid. p270).

Accordingly, and alongside the traditional quantitative methods of analysing consistency and differences in responses, listeners' graphical representations have been examined for ambiguities, outliers and anomalies. The aim of this more qualitative analysis was firstly to detect any unusual responses which suggested difficulties for listeners when graphically representing their experiences. Moreover, this identification process has been used to identify the existence of ambiguities which could prevent the researcher from understanding listeners' experiences through their graphically descriptions.

Since the revelation of ambiguities in either category could impede the attainment of the research objective, it was also considered prudent to continue the investigation of these anomalies and to provide answers to the following questions:

- Where are ambiguities occurring?
- Why may ambiguities be occurring?
- How can these ambiguities be minimised?
Thus, the systematic development and evaluation of the descriptive graphical language according to both a quantitative analysis and the identification and understanding of ambiguities and anomalies in listeners’ responses, should provide a clear indication of the extent to which the research objective has been fulfilled.

Accordingly, the development of the language should be able to provide answers to one further question:

- Within the context of this research\(^{31}\), what are the salient spatial attributes (experienced by listeners) that can be structured and represented using the developed graphical language?

Finally, further to establishing the effectiveness of the descriptive graphical language and identifying the salient spatial experiences that can be communicated graphically, it is appropriate to consider the external validity of any conclusions that can be drawn from the research. More specifically, since the descriptive graphical language is being developed to structure, represent and communicate listeners’ experiences, it is sensible to identify the listening population for whom the language can be considered an effective communication medium.

Even though, historically, subjective audio evaluation has favoured the use of experienced listeners when confirming the perceptual attributes of audio devices, highly experienced listeners may hinder the exploration and development of a descriptive graphical language because of their existing knowledge. Furthermore, the use of only experienced listeners will restrict the external validity of the language, as it cannot be known if a wider cross-section of the population will graphically describe their experiences in the same way as the expert listeners. Miles and Huberman actively encourage the inclusion of “extreme cases” (for example individuals known to have a strong bias in any direction) in any population sample as a means of “verifying and confirming conclusions” (ibid. p270). The authors believe that obtaining surprising responses from extreme cases (for example understandable graphical descriptions from participants with no previous listening experience) enables a greater degree of confidence to be placed in research conclusions. It has therefore been considered essential to investigate the complexities and diversities of the listening population during the development of the descriptive graphical language.

\(^{31}\) It is acknowledged that investigation settings, the choice of programme material, listening environment, and even the choice of listeners will influence the auditory spatial experiences that are considered salient.
1.6 **Original contributions of the author**

The purpose of this section is to briefly summarise how the author has contributed to subjective audio evaluation through the development of the descriptive graphical language. The originality of the work when placed within the context of contemporary and historical subjective audio evaluation can be demonstrated by:

- **The positioning of meaning with the listener**
  
  This work offers an alternative perspective to that favoured in more traditional subjective audio evaluations. Here, obtaining an understanding of the listeners' experiences (and not confirming the researcher's knowledge) of the external stimulus, provides the motivation for the work.

- **The use of elicited or developed graphical terminology**
  
  Although researchers have previously elicited verbal terminology from participants in subjective evaluation, the elicitation (or development) of a language of descriptive graphical terminology is novel.

- **An investigation of the perceived spatial characteristics of reproduced audio**
  
  The graphical language is used to describe spatial attributes of reproduced audio (as they are experienced by listeners). Although, with the continued development of multichannel reproduction systems, research in this area is becoming increasingly popular, perceived spatial qualities of reproduced sound remain less well documented than timbral or distortion artefacts.

- **The evolution of a novel method for developing descriptive languages**
  
  As a result of developing the descriptive graphical language, a novel language-development method has evolved for use in subjective audio evaluation.

The thesis will therefore provide details of the systematic development and evaluation of a descriptive graphical language for communicating listeners' experiences of spatial attributes in reproduced audio.

Although each of the following is allied to the current thesis, research has not been conducted into:

- **A listener's preference for particular spatial attributes**: The thesis concerns the graphical description of auditory spatial experiences.

- **A comparison of descriptive media**: The graphical language is proposed as an alternative or complementary (rather than superior) descriptive medium to those currently in use within subjective audio evaluation.

- **Any systematic correlation of subjective responses with objective measures**.
1.7 Thesis overview

The thesis is divided into three sections; an introduction, a development section and a presentation of the research conclusions.

Introduction (Chapter 1: Historical perspectives; research objective; evaluating the effectiveness of the research; original contributions of the author)

The task of this chapter has been to introduce the research. Chapter 1 commenced by outlining the rationale for the research project and continued by summarising the historical perspectives framing the development of the thesis. With the proposed research placed in appropriate context, the specific research objective guiding the development of a descriptive graphical language has been presented, and the means for evaluating the effectiveness of the research project outlined. The chapter has concluded by briefly stating the original contributions of the author to contemporary thinking and practice in subjective audio evaluation.

Development (Chapter 2: The development of a descriptive graphical language; a summary of three initial investigations)

Chapter 2 charts the early development of the research project. In its inaugural stages of development, listeners were involved in three investigations in which they described their auditory spatial experiences of reproduced audio stimuli using their own graphical descriptors. Chapter 2 contains summaries of these initial investigations, detailing procedures, notable observations from the analysis of listeners' responses and conclusions based on the findings of each study. Other than obtaining preliminary information about the auditory spatial experiences that could be represented graphically, the objective for each of the investigations was to identify ambiguities and limitations that could prevent the attainment of the research objective and for each subsequent investigation to improve on the insufficiencies of its predecessors. The chapter is written with an emphasis on the development of the descriptive graphical language with respect to the findings of the individual investigations.

Development (Chapter 3: Clarifying individual listener's descriptions; the development of a universal graphical language)

Following the initial investigations, it was determined that the ambiguities which were present in listeners' responses required more in-depth exploration. This decision was made by the researcher because it was essential to identify where and why these anomalies were occurring (and ultimately how they could be minimised) in order to improve the likelihood that the research objective would be fulfilled. In the opening part of chapter 3 (clarifying individual listeners' descriptions) a summary is presented of an investigation which elicited individual graphical descriptions from listeners and then returned these responses to the individual listeners to obtain a verbal clarification of what was being depicted. Using this method, ambiguities in the descriptive process were identified. Following the discussions with individual listeners, it was determined that one way of resolving the problematic
ambiguities in their responses would be to develop a language of descriptors which could be used by all listeners for the description of their individual auditory spatial experiences. The development of this universal graphical language (U-GAL) involved a process of inter-subjective discussion, whereby listeners agreed on the descriptors that would be included in their mutual language and the individual experiences that would be represented by these common graphical terms. Further details of this process are provided in the second part of chapter 3.

**Development (Chapter 4: Evaluating U-GAL)**

In this last chapter of the development section, details are provided concerning the formal evaluation of the universal language (developed by listeners and summarised in the second part of chapter 3). Chapter 4 presents an account of the evaluation process and a review of how individual listeners represented their experiences using the universal descriptors. An assessment of the suitability of the universal language for structuring, representing and communicating listeners' auditory spatial experiences is presented in this part of chapter 4 alongside an evaluation of whether the objective of the research project has been attained. Chapter 4 concludes by presenting a model of the routes through the descriptive process which are possible at this stage in the development of the descriptive graphical language.

**Conclusions (Chapter 5: Conclusions; further work)**

The thesis concludes with a brief section summarising the research project and outlining the author's original contributions to contemporary thinking and practice in subjective audio evaluation. The research conclusions are presented together with an acknowledgement that any statement regarding the effectiveness of the descriptive graphical language should be made with reference to the context in which this language was developed. A second section in chapter 5 outlines the further development proposed for the current descriptive language; a process which continues to employ the novel language-development method evolved during this research.

**Appendices**

The main body of the appendices (numbered from 2 - 4 to correspond with the development chapters of this thesis) contain illustrative accompanying information from the development and evaluation phases of the descriptive language. Within these appendices are copies of any instructions provided to listeners, exemplary graphical responses, transcripts of discussion sessions and graphs and tables from any statistical analyses.

A special observation needs to be made about the appendices. Since the majority of the work completed for this thesis is in a graphical format, the appendices for the development chapters (chapters 2, 3 and 4) are extensive. Consequently, the appendices have their own contents page, listing the tables and figures contained within its pages. The decision to include comprehensive appendices of graphical data may be criticised for being cumbersome. However, it was believed that the inclusion of
listeners' original graphical responses would offer the interested reader an opportunity for undertaking their own interpretation of the available data. Thus, appendices contain all graphical responses for listeners who are referred to in the main text. But, for purposes of brevity and immediate illustration, only selected responses from these appendices are included within the body of the thesis.
CHAPTER 2

The development of a descriptive graphical language
A summary of three initial investigations
2.0 Chapter overview

Chapter 2 provides a detailed summary of three separate investigations conducted in the early stages of the development and evaluation of the descriptive graphical language.

The first of the investigations to be detailed in section 2.2 of the chapter is a small scale pilot study. The first aim of this investigation was for the researcher to acquire an elementary understanding of the auditory spatial experiences that listeners could structure and represent graphically using their own descriptive graphical languages (GAL). A second aim was to ascertain how the graphical data elicited from listeners could be analysed. The section contains a review of the investigation’s procedure and provides a rudimentary analysis of listeners’ graphical data before highlighting limitations identified as a result of conducting this preliminary study.

Section 2.3 details how the second investigation addressed the inadequacies recognised in the pilot study, specifically with respect to the limited listening population, the complexity of the programme material and the investigation procedure itself. Improvements to these conditions are presented before a detailed analysis of listeners’ graphical descriptions is described. The section concludes with a summary of the investigation’s findings and proposals for further work.

Section 2.4 describes the last of the initial investigations. This third investigation once again sought to improve on limitations and ambiguous elements from earlier studies. Briefly, the listening population was expanded (to include a greater number of less experienced listeners), programme material altered, and differences between the physical stimuli used in this investigation minimised to ascertain whether the manipulation of these variables would have any influences on listeners’ graphical descriptions. An overview of the investigation procedure and a detailed graphical and statistical analysis of listeners’ responses is provided before acknowledgement is made of the limitations of the graphical language at this stage of the research.

In section 2.5 a summary outlines the principal findings from each study. The chapter then concludes by acknowledging the positive outcomes, limitations and future direction of the research at the end of the initial GAL investigation phase.
2.1 Introduction

To reiterate the objective of the research: the aim in developing a descriptive graphical language is to elaborate a medium for structuring and representing listeners' auditory spatial experiences, in order that - when communicated - these experiences may be understood by the researcher.

Fundamentally, the accomplishment of the research objective necessitates that listeners structure and represent their experiences in a graphical language. In chapter 1 (section 1.2.5) it was shown that, although an individual may find it difficult to express their experiences linguistically, the language system eventually constructed by that individual will have meaning for them; a certainty not necessarily associated with provided language. It was established in section 1.2.5 and later in 1.2.6 (when the work of Berg (2002) was discussed) that the use of an individually constructed language system enables the structuring of a listeners' auditory spatial experiences and the representation of their own personal constructs within the constraints of the chosen medium.

Because the research concentrated on the development of a descriptive graphical language (GAL) capable of structuring and representing listeners' experiences, it was resolved that listeners would develop their own individual graphical communication systems. It was envisaged that the development of distinct languages would provide individuals with a means of exploring and structuring their own auditory spatial experiences and enable participants in audio evaluation to bridge the gap between experience and representation. However, the use of an individual GAL by a listener would not automatically guarantee the successful accomplishment of the research objective. For although the use of an individual graphical language would ensure that a graphical representation was meaningful to the individual listener, the meaning being communicated through the graphical representation (a listener's auditory spatial experiences) needed to be understood by the researcher. Only when the researcher had an understanding of listeners' spatial experiences could the GAL be said to have successfully met the research objective. And, since it was not known whether individual graphical representations would be understandable to the researcher, initial investigation was required.

Accordingly, when the initial GAL evaluations were conducted, there were numerous questions to be addressed. Firstly, with listeners using their own individual graphical languages, the investigations provided the earliest opportunity for listeners to describe the spatial characteristics of reproduced audio as experienced. Thus, the studies provided the researcher with a possibility for acquiring an understanding of the salient spatial qualities of listeners' experiences that could be represented using a graphical medium. Correspondingly, these investigations gave the researcher an initial indication of whether a descriptive graphical language could be used by an individual as an effective communication medium. But what could be considered a measure of effectiveness in this context?

Since meaning originated with the listener, and the task of each listener's descriptive graphical language was to communicate a representation of this meaning to the researcher, it was not sufficient to determine the success of a GAL using a confirmatory analysis founded on the researcher's knowledge.
A summary of three initial investigations

Chapter 2 - Development

of an external stimulus. Essentially, even though it would be possible to obtain an indication of whether a GAL could be used consistently to describe differences in external stimuli, it was considered unwise to base a measure of GAL’s success entirely on an evaluation using quantitative criteria.

Consequently, and as outlined in the preceding chapter¹, the initial investigations provided the earliest opportunity for the effectiveness of the communication medium to be evaluated using qualitative processes. Miles and Huberman (1994) had proposed that the examination of any outliers within a data set (and an investigation of surprising results) would help build a better explanation of the data. Thus it was decided that - further to identifying consistency in listeners’ graphical language use and ascertaining whether listeners were representing differences in external stimuli - it would be prudent to examine listeners’ responses for ambiguities, outliers and anomalies to highlight any potential problems for the researcher when attempting to understand listeners’ experiences through their representations and to obtain a clearer understanding of what was being described. Whilst considering the effectiveness of the descriptive graphical language, the initial investigations also provided the opportunity for identifying the listening population for whom the language could be said to be effective and any limits to this population.

Further considerations for the initial investigations were made with respect to the practical aspects of conducting audio evaluations. For example, it was necessary to identify how to analyse listeners’ responses for similarities, differences and ambiguities. Furthermore, the investigation setting - the means of evaluating the graphical language - required deliberation; in particular, how would listeners’ graphical representations be elicited and what conditions would be manipulated within the studies to give rise to listeners’ experiences? Although the foundation for this thesis lay in the development and evaluation of the descriptive graphical language, it was essential to contemplate the audio devices that would be used in the initial investigations to provide the external stimuli for the listeners’ auditory spatial experiences; the eventual role for the developed language being the provision of a valid method for evaluating audio devices.

Thus, the initial investigations were required not only to identify the salient auditory spatial experiences that could be structured and represented using a graphical language, but to establish the practical means by which the success of a GAL could be evaluated.

¹ See practical considerations for a more qualitative understanding of listeners’ experiences in section 1.2.4 of this thesis and section 1.5.
2.2 A pilot study into the graphical description of spatial attributes

The aim of the pilot investigation was to establish the type of spatial information that could be described graphically using a listener’s own language and to identify methods for evaluating graphical data. The investigation was on a very small scale and used only three listeners, all of whom were musically competent students at the University of Surrey. The investigation required that participants graphically describe their experiences of the spatial characteristics of two-channel stereo programme material. The various musical extracts were reproduced over three loudspeaker pairs which were positioned behind an acoustically transparent curtain at three locations in a specially designed listening room. The listeners were themselves required to listen from three different locations within the room. The pilot investigation therefore looked at manipulating many of the variables considered influential in subjective audio evaluation - specifically loudspeaker location, listening location and programme material. Although using multiple variables added to the complexity of the investigation, a pilot study was believed to be the perfect opportunity to experiment. Thus, the greatest breadth of information could be elicited about participants’ experiences, and data could be obtained about how the manipulation of the various factors influenced listeners’ graphical descriptions.

2.2.1 Details of the investigation setting

Stimuli

The five musical extracts used as stimuli in the investigation were taken from commercially available compact disc (CD) recordings. Music is by its very nature a complex stimulus. As identified in section 1.2.2 of chapter 1, the use of complex programme material may cause problems for listeners when they attempt to describe their experiences as this time-variable stimulus makes it difficult for listeners to describe their experiences. However (even though the selected music was not simple with respect to instrument content or the complexity of the melodic line) it was decided that this initial investigation would be an opportunity to examine how participants would describe these complex but recognisable stimuli, and in particular to identify spatial attributes that could be graphically represented.

Musical items were chosen for their spatial characteristics. It was decided (by the researcher) that a combination of small and large ensembles would be used to obtain a variety of descriptions from listeners. Two small extracts were selected (a jazz and a blues ensemble) as the location of instruments within each of these ensembles was well defined within the stereo scene. An extract of choral music - a live recording consisting of a soloist, choir and organ in a church setting - was chosen to present listeners with a less well-defined stimulus and a sense of a larger environment. An orchestral ensemble with violin soloist was used for the spatial differences between soloist and the large orchestra. Finally a live stadium recording of a folk-rock ensemble was selected. This extract contained well defined instruments, a sense of space between instruments and distance cues; the sound of an audience applauding was present on the recording. A 90-second segment from each musical item was looped onto digital tape. The editing of the programme material required careful consideration in order that
A summary of three initial investigations

Chapter 2 - Development

Auditory spatial experiences be revealed by each musical extract. The principal instruments and spatial content of each stimulus - as defined verbally by the researcher - are outlined in Table 2.2.1.

Table 2.2.1 Information about the pilot investigation programme material

<table>
<thead>
<tr>
<th>Name</th>
<th>Principal and key instruments</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jazz</td>
<td>Trumpet, drums, piano, string bass</td>
<td>Well defined instruments (localisability or locatedness)</td>
</tr>
<tr>
<td>Blues</td>
<td>Male voice, piano, drums, guitar, bass</td>
<td>Well defined instruments</td>
</tr>
<tr>
<td>Choral</td>
<td>Female voice, choir, organ</td>
<td>Location &amp; size of ensemble and individual instruments, 'space'</td>
</tr>
<tr>
<td>Rock</td>
<td>Lead guitar, guitars, audience, percussion</td>
<td>Distance to ensemble, size of ensemble and instrument location</td>
</tr>
<tr>
<td>Classic</td>
<td>Solo violin, orchestra, (wind, horn, strings)</td>
<td>Location &amp; size of solo instrument and instrument groups</td>
</tr>
</tbody>
</table>

Loudspeaker and listening locations

Figure 2.2.1 Loudspeaker & listening locations

The three loudspeaker locations and listening positions employed in the pilot investigation are illustrated in figure 2.2.1. The first loudspeaker pair (the light grey pair depicted in the figure) was located at ear height and subtended an angle of ±30° around the central listening position (seat '2'). This pairing of loudspeaker and listening location represented the optimum for listening to two-channel stereo material. Other loudspeaker and listening locations were not so favourable for two-channel stereo listening. Yet the inclusion of these speakers was not completely arbitrary, since they were representative of the more practical domestic listening environment. The second loudspeaker pair (figure 2.2.1, black) was positioned at the wider angle of ±60° around the central listening location and the final loudspeakers (dark grey) were positioned in the rear corners of the listening space and elevated to position the tweeters at 1.65m above the ground. Loudspeakers were level-aligned to ensure that the loudness of their output was equal for all pairs at all locations. Although listeners were not comparing and evaluating the loudspeakers themselves, this level alignment minimised the opportunity for listeners' auditory spatial experiences to be influenced by differences in loudness.

For the listening locations, both sub-optimal positions (seats 1 and 3) were situated one metre away from seat 2, with seat 3 located one metre further back from the other seats. The different listening and loudspeaker locations were ultimately chosen to identify whether listeners' graphical descriptions of their auditory spatial experiences would be influenced by their own listening position and the locations of the (hidden) loudspeakers.

Investigation procedure

The graphical investigation was part of a larger informal study of elicitation methods. Prior to their involvement with the graphical investigation, the three listeners had participated in a verbal elicitation stage. In this earlier stage, listeners were presented with triads of musical stimuli (the same as
employed in the graphical investigation) and - using a ‘repertory grid’ based technique - were asked to describe how each stimulus was similar to or different from the remaining stimuli with respect to the spatial attributes experienced. As all three listeners had been part of this verbal phase, no definition of what constituted a spatial characteristic was provided for listeners when they came to participate in the graphical elicitation. Rather, listeners were only instructed to draw what they perceived to be the spatial attributes of the reproduced stimuli.

The graphical investigation required the three listeners to participate in seven sessions or runs. These runs allowed for each musical stimulus to be reproduced over one pair of loudspeakers and for two of the extracts to be presented a second time. However, only one of these presentations may be considered a repeat, since the other reproduced the same stimulus item over a different loudspeaker pair in a different location. Listeners were not aware of the loudspeakers reproducing the material or their location, as speakers were positioned behind the (acoustically transparent) curtain. During each run all three participants were present and heard the same extract of music. Additionally, within each run, the listeners were required to change their listening location to include each of the three seats. The same programme material was repeated during each run until all participants had described the spatial characteristics of the stimulus from all listening locations.

The participants were instructed to record their descriptions using the pencils and sheets of tracing paper provided. Although the tracing paper was blank, an A4 sized guide sheet (figure 2.2.2) was provided which illustrated the three listening locations, the sides and perceived centre of the room, and the acoustically transparent curtain (the thick black line in figure 2.2.2). Participants were required to mark their current listening position with a cross and draw all guiding lines onto their tracing paper at the start of each run. Guiding lines were provided for listeners as it was established (section 1.2.7 of this thesis) that individuals use a system of fixed landmarks to co-ordinate their orientation within space. Listeners were therefore provided with this information to aid in the two-dimensional representation of their three-dimensional auditory spatial experiences.

2 A method introduced in section 1.2.4 of this thesis
3 It should be noted that loudspeaker pairs (although aligned for loudness) were from different manufacturers and could therefore be perceived as having different spatial (and timbral) characteristics.
Although listeners were familiar with the programme material and verbal elicitation, when asked to graphically represent the spatial attributes of these reproduced extracts on the tracing paper response sheets, listeners were initially bemused because they were not sure how to draw sound. In answer to queries about how to depict their spatial experiences, listeners were informed to use any appropriate graphical style - since the aim was for them to develop their own languages of graphical descriptors. Following further quizzical expressions, a rough sketch (figure 2.2.3) was produced for the listeners to identify a potential means of graphically representing their experiences. However, listeners were still encouraged to develop their own representation style. Furthermore, although a potential graphical response style was (reluctantly) provided, listeners were not issued with any information about the spatial attributes they could draw, since these auditory spatial experiences had to originate from the listener and not from the researcher. The task for the listener was completed when they had provided graphical descriptions of each stimulus from all three listening locations. Depictions were analysed by the researcher to obtain a preliminary understanding of the spatial characteristics of listeners' experiences that could be represented graphically.

2.2.2 Analysis of responses

Listeners responses were analysed using three different techniques. Firstly the depictions were studied individually to see if any trends could be established. Secondly - by overlaying the responses of the different participants to the same stimulus - plots were created which indicated whether there were any similarities between listeners' descriptions. And finally, using the graphical analysis of listeners' individual and group descriptions as a guide, a selection of spatial attributes was identified and a numerical measure of each of these graphical attributes was derived for further analysis.

Analysing listeners' responses

By studying listeners' graphical descriptions, it was immediately obvious that each had employed a slightly different drawing style, and that some similarities remained between listeners' descriptions and the example provided by the researcher. Representation styles ranged from very simple to relatively detailed. Figures 2.2.4 – 2.2.6 depict the three listeners' graphical responses for the folk rock stimulus when they sat at the central listening location (seat 2) and the music was reproduced over the loudspeaker pair in the optimal location.
Figure 2.2.4 is representative of the simple style used throughout by listener 1. At all times, other than when a specific instrument (for instance the guitar in the above depiction) was notable, listener 1 chose to use a single rectangular image to represent the attributes of the ensemble. In contrast, both other listeners chose more complex methods of depiction, where spatial characteristics of individual instruments within each ensemble were described independently. Like listener 1, listener 2 (figure 2.2.5) used rectangles to depict his experiences, whilst listener 3 (figure 2.2.6) chose to represent the different instruments, and the audience heard on the recording, using circles and curves.

The most noticeable difference between listeners when depicting the various musical extracts was in the number of instruments represented. Two possible reasons were proposed for these differences. Firstly, instruments could have been omitted from the graphical descriptions as a result of listener uncertainty. For example, a listener could have avoided depicting an instrument when it was not known where it was, what it was, or how to represent it graphically. An alternative reason for these differences was provided by the complexity of the musical stimuli. Because each stimulus contained many instruments, it was feasible that listeners only focused on describing selected instruments within entire ensembles - those they believed to be important. After a brief inspection of their responses, it was possible to confirm that all listeners had indeed depicted the same select instruments for each ensemble (or at least they were labelling some instruments with the same verbal label). The instruments most commonly selected for inclusion in listeners' graphical representations were the principal instruments within each ensemble (for example, the trumpet in the jazz ensemble, the voice in the blues ensemble and the lead guitar from the folk rock ensemble) and their analysis enabled further information to be gleaned from listeners' responses.

Plots were constructed of the principal instrument in each ensemble as represented by all listeners. As depictions were recorded on tracing paper, the creation of principal instrument plots involved lining up the cues provided by the guide sheet and re-drawing all listeners' responses, from all listening locations4, on a single piece of paper. Exemplary principal instrument plots are provided for the folk rock music (figure 2.2.7) and the voice from the blues ensemble (figure 2.2.8) - which were both reproduced over the optimal loudspeaker pair - and for the jazz extract (2.2.9) which used the rear loudspeaker pair. An examination of principal instrument plots suggested that listeners' graphical responses were not as random as first believed, since listeners all appeared to be placing the same instruments at similar locations behind the curtain.

Similarities in listeners' representations of principal instruments suggested that ambiguities in their responses to entire ensembles could be reduced if the selected programme material was less complex. Support for this assertion occurred in the form of listeners' responses to the simplest music (in terms of instrument number) used in the pilot investigation.

4 Numbers on each plot refer to the seat from which that principal instrument was depicted.
Within the choral ensemble, only three groups of instruments were present; a choir, vocal soloist and organ. When this ensemble was reproduced over the loudspeakers in the sub-optimal ±60° location, the same instruments were represented by the listeners\(^5\), as illustrated in figures 2.2.10 and 2.2.11 (listeners’ representations from the central listening location). Furthermore, when descriptions of the vocal soloist from this ensemble were plotted together for both listeners and for each listening location (see figure 2.2.12), a degree of correlation was visible in the location and general size of this source. Consequently, one initial conclusion of the pilot investigation was that the simplification of programme material would make the task of the listener less complex and their responses easier to understand. A second conclusion was that listeners’ graphical descriptions were being influenced by their listening location and that of the loudspeakers.

Figure 2.2.12 provides a principal instrument plot of the vocal soloist from the choral ensemble when reproduced over loudspeakers in the sub-optimal ±60° (around centre) location. When listeners described the soloist from seat 1, graphical descriptions were skewed to the left of centre, yet the same soloist was described to the right of the graphical response sheet when listeners sat in either seat 2 or seat 3. Differences in location were less visible when the lead guitar (illustrated earlier in figure 2.2.7) and the voice from the blues ensemble (figure 2.2.8) were described, with both instruments depicted as more or less central from all three seats. The implication was that loudspeaker location (earlier

\(^5\) Only two listeners remained at this stage of the graphical investigation, the third having to leave due to other commitments.
ensembles had been reproduced over the optimally located loudspeaker pair) was an influence on listeners' graphical descriptions when combined with listening location. This interaction was more fully illustrated using 'total image' plots.

**Analysing total image plots**

After briefly examining individual listeners' responses and further clarifying these responses using principal instrument plots, a set of total image plots were created. Essentially, each total image plot provided an overview of all the graphical information that existed for the same stimulus. Within each plot all instruments from the same musical ensemble were depicted together, as described by all listeners from all listening locations. As with principal instrument plots, individual listeners were not identified (other than by their depiction style) and numbers on graphical descriptors within each plot provided a reference of the seat from which descriptions was elicited.

Although too complex to study in any detail (due to differences in the number of instruments being represented), what was again obvious from the total image plots was the influence of the participants' listening location on their graphical description. When the plot of the folk rock ensemble (figure 2.2.13) was examined, listeners' responses from all three listening locations (seats 1, 2, and 3) were clustered around the centre of the response sheet. Since details were not known about how the programme material was recorded, it was impossible to state with certainty that the stability in listeners' responses (the similarity in descriptions from all listening locations) resulted from the use of the optimally located loudspeakers. However, this was a definite possibility, as listening location was more influential when the jazz ensemble was reproduced over the loudspeakers at the rear of the listening room (figure 2.2.14) and when the same ensemble used the loudspeakers positioned at ±60° around the central listening location (figure 2.2.15). In both these sub-optimal situations, listeners depicted the ensembles towards the edges of the response sheet when they were sat in either of the sub-optimal listening locations (seats 1 and 3). The skew of listeners' descriptions was particularly noticeable when listeners sat in seat 1 and were played the jazz ensemble over loudspeakers located at ±60° (figure 2.2.15). It was likely (due to the precedence effect6) that the increased skew, occurred as a result of seat 1's proximity to the left speaker.

---

6 The precedence effect can be briefly summarised as the influence of time on the perceived location of a sound. A sound's perceived location will alter according to the proximity of a listener to the loudspeakers reproducing that sound. If the listener is equidistant from a pair of loudspeakers, the reproduced sound will appear as intended. However, with a listener positioned closer to one loudspeaker, sound will arrive earlier at the listener's ears from this origin and the sound's perceived location will shift towards the nearest speaker. (see Rumsey, 2001, p26)
Numerical data analysis

The analysis of graphical descriptions had identified the need for simpler programme material and had illustrated how listeners' representations appeared to be influenced by their listening location and that of the loudspeakers. It was decided that further information might be extracted from listeners' responses if these were translated into numerical measurements and an uncomplicated data analysis conducted on these measurements. Because (with the exception of the very simplest stimuli) listeners had chosen to represent different instruments within each ensemble - and since this was only a rudimentary numerical analysis - measurements concentrated on the spatial characteristics of the whole ensemble rather than the individual instruments within each ensemble. The choice of which spatial characteristics to measure was determined from the analysis of listeners' graphical descriptions.

Differences in where listeners were locating the various ensembles had already been indicated in the analysis of their graphical depictions. Thus, for each musical extract it appeared feasible to obtain a numerical measure of the amount of skew\(^7\) associated with a listener's description when their listening position altered. Skew was calculated as an angle. The vertical line on the graphical response form (illustrated by the thick dotted line in figure 2.2.16) assumed the value of 0° and the skew angle was

\(^7\) For the purpose of this investigation, skew refers to the amount of shift a description undergoes away from the central line on the graphical response sheet.
calculated by drawing a line (the thick unbroken line in figure 2.2.16) between 0° and the centre of the depicted ensemble.

Several other spatial attributes were considered for this rudimentary numerical analysis. In addition to the graphical examination of listeners' responses, the decision about whether or not to measure a spatial attribute was taken with respect to known verbal descriptors of auditory space. Based on this combined evidence it was finally decided that two further spatial attributes should be measured, specifically ensemble width and ensemble depth; numbers 1 and 2 on figure 2.2.17 respectively. Ensemble depth appeared to differ from listener to listener when describing some ensembles (see earlier figures 2.2.10 and 2.2.11) whilst the width of an ensemble seemed to be influenced by listening location. It was therefore interesting to identify what a numerical analysis could reveal about listeners' descriptions of these attributes.

For ensemble width and depth, the conversion from graphical response to numerical data was uncomplicated and achieved by measuring (in millimetres) the distance between the furthest limits of a listener's depiction of an ensemble from left-to-right and front-to-back, illustrated previously in figure 2.2.17. Once measures were obtained for all listeners in all listening locations, it was possible to establish trends in the data in terms of the width and depth of listeners' descriptions. A statistical analysis was not conducted due to the limited size of the data set and the basic requirement of the analysis; to provide an overview of the numerical information that could be extracted from graphical descriptions. To maintain the simplicity of the numerical analysis, mean width and depth values were calculated for listeners' responses to the different stimuli as described from the three listening locations. Since only two listeners completed the full set of runs, the numerical analysis is based solely on the responses for these listeners.

When listeners described the width of the different ensembles from the various listening locations, the jazz extract (reproduced over loudspeakers in the rear corners of the listening room) emerged as the narrowest depiction from the central (seat 2) and rear (seat 3) listening locations. The same musical extract - reproduced over the loudspeakers positioned at ±60° around the central listening location -
was described as narrowest from seat 1. As indicated in figure 2.2.18, listeners described the folk rock music (reproduced over loudspeakers in the optimal locations) as widest for all listening locations.

When these width measures were initially considered, it appeared that listeners might have been influenced by their knowledge of how the various ensembles should have sounded - with physically larger ensembles (for example the choral work in the reverberant environment and the live recording of the folk rock) being depicted as wider than the smaller drier recordings of the jazz and blues ensembles. However, this observation was not as straightforward as first thought, as listening location also appeared to influence the width of listeners’ representations; a finding that was anticipated by the previous graphical analysis of listeners’ descriptions. In general, representations from seat 2 measured wider than depictions from the two sub-optimal listening locations and, with the exception of the classical music reproduced over the rear loudspeakers, the narrowest depictions occurred when listeners were in seat 1. As seat 1 was situated nearest to the acoustically transparent curtain on the left of the listening space, this response was a possible consequence of the listeners’ proximity to the optimal and sub-optimal (±60° around centre) loudspeaker locations.

In addition to ensemble width, the distance between front and rear most points of each listener’s graphical descriptions were measured to provide an indication of how listeners were representing the depth of the various ensembles. However, unlike for the measurements of ensemble width, measurements for ensemble depth did not produce much in the way of easily understandable data.

Figure 2.2.19 All listeners descriptions of ensemble depth (measured in mm)

When mean measurements of listeners’ descriptions of ensemble depth were examined (see figure 2.2.19), few trends were noticeable. The most obvious finding was that listeners depicted the choral ensemble as deepest from all three listening locations. The deeper portrayal of the choral music could be attributed to the influence of one listener (listener 2, figure 2.2.10), who consistently described the organ in this ensemble at the rear of the listening space. The jazz extract was described as shallowest from seats 1 and 3 when this was reproduced over the rear loudspeakers. The classical music, reproduced from the same location for the first time, was described as shallowest from the central listening location. However, when described for a second time, the classical music was one of the deepest. Not only were there unexplainable differences in how listeners were describing the same stimulus but little sense could be made of listeners descriptions according to their listening location, with no individual seat associated with notably deeper depictions.
The skew associated with the graphical description of each ensemble was converted into degrees using the procedure previously outlined in figure 2.2.16. The numerical analysis of these measurements indicated that (as predicted by an examination of listeners' descriptions) ensemble skew was associated with listening location. Mean measurements for ensemble skew were plotted in a bar chart (see figure 2.2.20), where negative values indicated a shift in listeners' depictions to the left of centre and positive values a skew to the right. When listeners sat in seat 1, descriptions of all stimuli were consistently skewed to the left of centre. Conversely, and as expected, listeners described the musical extracts to the right of centre from seat 3. And, when sat in the optimal location (seat 2), a less obvious skew was depicted in any particular direction with all measurements clustered around the central axis (0°).

Figure 2.2.20 All listeners descriptions of ensemble skew (measured in degrees)

2.2.3 Summary of findings

The aim of this pilot investigation was to explore the potential of a descriptive graphical language for representing listeners' experiences of spatial attributes in reproduced audio. A combination of graphical and rudimentary numerical analyses identified that listeners were able to describe differences in the location (skew) and width of complex musical ensembles from various listening locations when this material was reproduced from different loudspeaker locations. However, differences in listeners' descriptions of other spatial characteristics (in particular ensemble depth) could not be explained by the manipulation of variables in this study.

It is possible that listeners did not consider ensemble depth when describing their experiences of the various ensembles. When, in the verbal part of this pilot investigation, participants were asked to describe what they had heard, depth was not mentioned as an attribute of the reproduced sound. The selected musical extracts were full of spatial information, and since listeners were not asked to describe specific spatial characteristics, it is likely that they would have represented those attributes which were more pertinent to their experiences. Accordingly, if depth was not considered a primary spatial descriptor of two-channel stereo stimuli, the programme material's complexity may have added to the ambiguities in depth description, with listeners focusing their attention elsewhere in the complex...
material. The lack of coherent depth description could therefore have reflected listeners’ individual experiences, but it was equally possible that the response sheet may have caused problems for listeners when attempting to represent any front-back information. Whereas listeners could gauge the width of the room by acknowledging the position of the side walls, the rear wall of the listening space was obscured by the acoustically transparent curtain. The removal of this visual cue could have prevented participants from effectively scaling descriptions in this dimension. Whatever the reason, it appeared that several factors needed careful consideration before any further investigation of a descriptive graphical language.

Considerations for future graphical language investigations—listening population

One obvious limitation which needed rectifying in subsequent investigations was the number of participants. Because only three listeners were used in the pilot study (and only two of these completed descriptions of all musical extracts) any understanding of auditory spatial experiences that could be graphically depicted was limited and elementary. In particular, the use of such a small (and specific) set of listeners was problematic for the external validity of the results. Expressly, it could not be known whether the auditory spatial experiences described by this small sample of listeners were representative of a wider listening population.

Considerations for future graphical language investigations—programme material

A second cause for concern was the programme material. Rumsey (1998) had previously noted how the complexity of material could prevent listeners from identifying what they were evaluating. As listeners in the pilot investigation responded better (representing the same instruments as one another) when the reproduced extracts contained fewer instruments, it appeared that complexity may indeed have caused problems for the listeners. A further concern when it came to the choice of stimuli was identified in chapter 1 (section 1.2.2). Here, the use of commercially available material was criticised for having unknown technical properties which could prevent the researcher from attributing effects to specified causes. Since the aim of the initial (and any subsequent) investigation was for listeners to describe their experiences rather than the properties of an external stimulus, as known by the researcher, this lack of knowledge did not present too great a problem. However, with the use of commercially available material came the potential for listeners to describe their experiences based on their existing knowledge of how a stimulus should sound, rather than their current auditory experiences, and this could present problems for the validity of the work.

Thus, for subsequent studies, the simplification of the programme material and the use of created rather than existing material was advocated. Since the focus of the research was one of obtaining an understanding of how listeners experienced the spatial attributes of reproduced audio through their graphical descriptions, the simplicity of the programme material needed to be balanced with its effectiveness for obtaining a representation of these experiences. From the outset, any type of noise

---

10 Depth is considered in more detail in section 3.3 of chapter 3.
11 All listeners were students of sound recording at the University of Surrey
A summary of three initial investigations

was discounted as a suitable stimulus. Although noise was considered a simple stimulus - and one with no obvious spatial associations for the listener - noise was not typical of reproduced audio. It was finally decided that musical material would be created with limited numbers of instruments located in various positions within unconventional ensembles. It was believed that the creation of these ensembles would maintain the validity of the material whilst minimising its complexity. Furthermore, by placing instruments in different locations within unconventional ensembles, it was considered less likely that listeners would describe their existing knowledge rather than their current experiences. Consequently, with the listener unaware of the content of each stimulus prior to describing it, demand characteristics\(^{12}\) were minimised.

Considerations for future graphical language investigations – investigation procedure

The current investigation consisted of a limited number of runs which resulted in listeners only describing one musical extract twice. Although listeners' experiences may differ (quite validly) over time, the inclusion of repeats in investigations is advantageous because the presence of consistency in repeated responses can suggest where a graphical medium is being used reliably. Consequently, it is considered important for subsequent investigations to contain a greater number of repeated runs. A further improvement to the investigation procedure is in the scaling of the response sheet. It was believed that accurate scaling would avoid any confusion which could be caused by the visual landmarks on the response sheet not matching listeners' experiences when listeners translated between these three-dimensional egocentric experiences and the two-dimensional plan representation.

The aim of the pilot investigation was to establish the type of spatial information that could be described graphically using a listener's own descriptive language and to identify methods for evaluating graphical data. Even though results from the investigation were interesting, suggesting that listeners were able to describe some auditory spatial experiences using a graphical language - and that different analysis techniques could be used by the researcher when determining listeners' experiences through their graphical representation - the elementary nature of the study limited any conclusions which could be drawn. Thus, the only true conclusion at this stage of the development of a descriptive graphical language was the necessity for further research to address the many limitations of this initial study.

\(^{12}\) Any cue that may reveal the experimental hypothesis (See Coolican, 1996, p33)
2.3 A formal study of graphical description

The second investigation built on the work of the pilot study and, in particular, explored areas where limitations had previously been identified. Briefly, the 12 listeners involved in the study were asked to provide graphical descriptions of their auditory spatial experiences when musical programme material was reproduced over a multichannel audio system in a stationary vehicle. Employing their own descriptive graphical styles and paper response sheets, the listeners depicted selected spatial qualities as perceived from three different listening locations within the vehicle. Individual characteristics of listeners' depictions were examined before a statistical analysis was conducted using numerical measurements taken from these depictions. Further details of the investigation setting and procedure are outlined in the following section.

2.3.1 Details of the investigation setting

Listening location

Since listeners in the pilot investigation had demonstrated that a graphical language appeared capable of describing the auditory experience of how a reproduced scene skewed according to a listener's location, it was decided that participants would once again be asked to listen from different positions. However, rather than listening within a room, the investigation was conducted within a stationary vehicle as, in this environment, the driver and front-seat passenger would necessarily be in off-centre locations; sub-optimal for conventional stereo listening. As a result of the precedence effect (detailed earlier, but briefly, where a listener experiences a reproduced image steered towards the nearest loudspeaker) it was of interest to identify whether listeners would represent any differences in the spatial characteristics of programme material when sat at various positions within the vehicle. Three listening locations were investigated, namely the front left (driving seat), the rear centre location (for a more optimal, if less practical automotive listening location) and rear left location (to obtain information from a rear sub-optimal listening position).

Audio reproduction system

The influence of a second variable - the reproduction system within the vehicle - was also of interest in this investigation. The pilot investigation had identified how listeners' descriptions were less likely to indicate a skew in the reproduced scene (when combined with listening location) when the programme material was reproduced over loudspeakers in an optimal location for two-channel stereo listening. Although loudspeakers within the vehicle were not moveable, the same stereo programme material could be reproduced in various ways over the vehicle's audio system.

The decision to set all subsequent investigations within stationary vehicles was taken as a result of the interest shown by an audio manufacturer in the embryonic graphical language when a paper detailing the pilot investigation was presented at the Audio Engineering Society's 110th Convention (see List of Publications, Ford et al. 2001). Since the research objective centred around establishing the suitability of the descriptive graphical language for enabling the structuring, representation and communication of listeners' auditory spatial experiences, the move from listening room to vehicle was not considered detrimental, indeed it offered the opportunity to conduct further investigation into graphical description of listeners' experiences from off-centre listening locations.
The reproduction of spatial information proved a particular problem for automotive audio because of the lack of reverberation within the confined listening environment. The multichannel audio system employed throughout this investigation was developed to "enhance" the spatial reproduction of two-channel programme material within vehicles (Nind, 2001). Amongst the objectives for developing the audio system were the provision of "a wide, accurate stage and true sense of envelopment for each seat" (ibid. p5). One of the attributes of the audio system was to move the perceived location of reproduced sources away from the loudspeakers, in order that listeners would be "enveloped by the soundfield" and occupy "the same acoustic space as the material" (ibid. p7) rather than experience the sound as coming from the speaker closest to them. This effect was desirable within vehicles for "creating a larger listening space and a sweet spot\(^4\) that covers all seats" (ibid.).

For this investigation, the input for the multichannel audio system was a two-channel stereo signal, provided by a conventional CD of programme material. In one system setting, this signal was processed to derive different signals for each of the playback channels feeding the seven loudspeakers, located as illustrated in figure 2.3.1

![Figure 2.3.1 Architecture of the audio reproduction system within the vehicle](image)

The default setting for the multichannel reproduction system - illustrated in figure 2.3.2, and referred to in more detail by Nind (ibid.) - was used in this investigation. To obtain an output signal at the centre loudspeaker (C), the left and right input signals from the two-channel stereo material were summed and each attenuated by 7.5dB (preventing any narrowing of the reproduced image) before being fed to this loudspeaker. The front left and right signals (reproduced over LF and RF loudspeakers) were then filtered and delayed to provide signals for the side channels (LS and RS) and further delayed and filtered for the rear channels (LR and RR). Nind believed this process created an ambient soundfield and the perception of natural envelopment.

Further to its spatially enhanced setting (hereafter referred to as the processed setting), the audio system was capable of reproducing two-channel material in its original format or stereo setting. With the audio system in a stereo setting, six loudspeakers within the vehicle were used (all except 'C' in

\(^4\) Optimal listening area
A summary of three initial investigations

Chapter 2 - Development

Figure 2.3.1) and no spatial enhancement was made to the material being reproduced over these loudspeakers. As a result of the precedence effect (described earlier in this chapter) it was believed that for off-centre listening, spatial information would be compromised for material reproduced using this setting and listeners would describe stimuli towards their nearest door speaker. It was envisaged that the influence of the off-centre listening locations would be minimised when the audio system was used in its processed setting. Thus, the potential existed for listeners to describe differences in their auditory spatial experiences according to the setting used by the reproduction system.

Programme material

At the end of the pilot investigation, observations were made about the type of programme material that would be used in subsequent investigations. Ultimately it was decided that material would be kept simple - to ease the task of each listener when describing their experiences - and use real (as opposed to synthetically generated) instruments playing random phrases of music in unconventional combinations and locations. By creating unique ensembles it was believed that little precedent would exist in current repertoire, reducing the likelihood of listeners basing graphical descriptions on their existing knowledge of how an ensemble should sound, rather than their current experiences.

For the investigation, three items of musical stimuli were created using trios of instruments. The original instrumental extracts were taken from a CD of anechoic monaural recordings used in the Archimedes project (Hansen and Munch, 1991). The recordings chosen from this CD were a cello playing a sustained passage, a female voice repeating text in Danish, and an extract of drums (percussion). The decision to use the three instruments was taken because of their timbral diversity (each was distinct enough to be audible in the collective ensemble) and more critically each was selected for its spatial attributes, ranging from the predominantly low frequency content and continuous nature of the cello music to the transient higher frequencies of the percussive extract. Three trio ensembles were created by amplitude panning 30-second extracts of the individual mono recordings to different locations within a two-channel stereo scene as indicated by table 2.3.1. To create more realism in the stereo scenes, an equal amount of artificial reverberation was added to each ensemble.

15 In the stereo setting, the same two-channel programme material was reproduced in parallel left + right pairs throughout the vehicle. The only alteration was for the signal to be attenuated in the loudspeaker pairs towards the rear of the vehicle.

16 An anechoic environment is an artificially dry environment in which naturally occurring reflections are reduced to a minimum (see Rumsey, 2001, p2)

17 The extracts were taken from anechoic recordings and were consequently not typical of the two-channel stereo material available to listeners.

18 Programme material was edited by Russell Mason at the University of Surrey.
Programme material was recorded onto six CDs, each with its own running order. By manipulating the presentation pattern of the various extracts, the likelihood of listeners' descriptions reflecting a specific sequence in the presentation of this material was reduced.

**Participants**

One of the major limitations of the pilot study was the number of participants. Accordingly, for this investigation it was decided that the number of listeners should be increased from three to 12. Although not a substantial increase in the listening population, the decision to involve 12 participants was taken with reference to section 1.2.3 of chapter 1 in this thesis. As noted in this earlier section, the number of listeners required within an investigation was correlated with their level of experience; the higher the expertise or previous experience of the listener, the lower the quantity required to obtain reliable responses. Since the 12 listeners selected to participate in this study all had an active interest in music and/or audio engineering - and a precedent had been set by Koivuniemi and Zacharov (2001) for using 12 experienced listeners in a descriptive study - it was decided that this number would be sufficient to obtain a selection of spatial descriptions from listeners. However, by using only experienced (and in some cases highly trained) listeners, acknowledgement was made about the potential for their responses to be unrepresentative of a more generic listening population.

**Spatial attributes**

Participants in the pilot investigation had been presented with complex musical extracts and received no instruction about the spatial attributes to describe. This lack of restriction enabled a useful insight to be gained into the type of graphical descriptions that could be elicited from listeners. However, the sheer amount of graphical data provided by each participant - and differences in the number of instruments being represented - proved problematic when identifying the spatial characteristics that were being described. Accordingly, the analysis of listeners' depictions was rudimentary since it was not known where attention should focus to obtain a fuller understanding of listeners' experiences. It was therefore decided that this investigation would establish the characteristics participants were required to depict prior to its initiation.

Since listeners in the pilot investigation had described differences in the location of ensembles\(^\text{19}\) when sat in sub-optimal listening location - and since listeners in this investigation would be asked to listen

---

\(^{19}\) Ensemble location refers to the amount of shift a reproduced ensemble undergoes away from a designated reference position. For the purpose of this investigation, the reference position is taken as a line down the centre of the vehicle from front to back.

<table>
<thead>
<tr>
<th>Name</th>
<th>Location 1</th>
<th>Location 2</th>
<th>Location 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>percussion</td>
<td>Cello</td>
<td>Voice</td>
</tr>
<tr>
<td>2</td>
<td>voice</td>
<td>Percussion</td>
<td>cello</td>
</tr>
<tr>
<td>3</td>
<td>cello</td>
<td>Voice</td>
<td>Percussion</td>
</tr>
</tbody>
</table>
from three different positions in the stationary vehicle - this location characteristics was the first to be selected for further examination. The second attribute to be selected as a focus for this study was ensemble width\textsuperscript{20}. When responses from listeners in the pilot investigation were numerically analysed, differences had been identified between their descriptions of the width of various ensembles. Once again, this second investigation provided an opportunity for the width (in addition to the location) of reproduced material to be manipulated using the two settings of the vehicle's audio system. It was believed that the width of reproduced material could be increased when listeners listened to the stimuli over the processed system setting from sub-optimal listening locations. Consequently it was of interest to identify whether listeners' graphical descriptions would reflect these changes.

Restricting listeners to describing the ensemble width and location (skew from centre) of the various stimuli was believed beneficial in focusing listeners' attention and obtaining a clearer indication of how these specific characteristics were experienced and represented. However, it was acknowledged that this constraint could result in more pertinent auditory spatial experiences being overlooked by the investigation.

**Investigation procedure**

Listeners were asked to participate in nine runs which enabled the three programme items to be reproduced using both audio system settings (stereo and processed) and also allowed for the presentation of three repeat runs. Repeats were limited because of the unknown time element. Specifically, this was the first formal investigation of the graphical descriptive language and it was not known how much time individual listeners would require in order to describe their experiences and complete the investigation. However, the inclusion of selected repeats went some way towards remedying the inadequacy of this factor in the pilot study. Repeats were included to establish individual listener consistency. Although a listener could (quite validly) change their experiences over time - and consequently a lack of consistency did not necessarily suggest a failing of either the graphical language or a listener's ability to describe their experiences - where listeners' descriptions were consistent, it was possible to suggest an element of reliability in their language use.

Listeners participated individually in the descriptive task. During each run, listeners were asked to complete three (A3) response sheets, one for each of the listening locations; front left (driver's seat), rear centre, and rear left. As illustrated in figure 2.3.3, the two-dimensional response form depicted a scaled representation of the listening environment (the passenger compartment of the car) and included visual cues to improve the translation from egocentric experience to external response. Space was left on the sheet so that listeners could also depict experiences outside of the vehicle should this be required.

\textsuperscript{20} Ensemble width can be defined as a measure of the distance between the furthest extremities of an ensemble of instruments (from left-to-right) as graphically depicted by a listener.
Before each run, the listener was given three response sheets with each sheet crossed at one of the three listening locations to indicate the seat order for the run. Listening location order was randomised to avoid listeners always starting a run in the same seat. During each of the nine runs, the CD was placed on track-repeat and the computer controlled reproduction system was set to the required mode (either stereo or processed) by the researcher whilst out of sight of the listener. Listeners were informed that there were no correct answers - the objective was merely for them to describe their experiences - and to spend as long as necessary completing the response sheets. Listeners could take a break whenever tired, which was usually around run seven.

A day prior to their participation in the investigation, listeners had been presented with written instructions (see appendix 2.A) containing full details of their task. Within these instruction were verbal definitions of ensemble width and location; the spatial attributes listeners were requested to focus their attention on. Unlike with the pilot investigation, listeners were not provided with any form of exemplary graphical description prior to their participation, and were told to draw their experiences of the two attributes using any depiction style. Since all listeners were new to graphical representation, it was common (at the end of the first run) for listeners to seek the researcher’s confirmation that their chosen drawing style was adequate; listeners were always informed that their graphical descriptions were fine so long as they enabled the description of their experiences.

2.3.2 Analysis of graphical descriptions
Listeners' graphical descriptions were analysed using two methods. Firstly, individual responses were examined to establish trends in how listeners were describing the width and location of ensembles reproduced over the different audio settings and to identify when listeners were being consistent in their graphical description. Individual responses were then measured to enable the confirmation of any statistically significant differences between descriptions. Although listeners were not required to describe differences between stimuli when not experienced, the physical differences in listening location and reproduction system could have had an influence on listeners’ experiences and it was therefore of interest to see whether this would be reflected in their depictions.

Analysis of individual responses
Graphical descriptions to accompany this analysis are presented in appendix 2.B\(^2\). When these individual descriptions were examined, one visible improvement over the pilot investigation was that all listeners had chosen to describe the same trio of instruments in their responses, suggesting that the programme material was of a sufficiently simple nature to enable its description. It was also recognised

\(^2\) For ease of reference, specific figures will be included in the main text. However, readers are directed to appendix 2.B for a more extensive selection of listeners' responses
that, as with listeners in the pilot investigation, this more extensive group of listeners (who had received no instruction about how to draw their experiences) predominantly used rectangles or circular graphical descriptors, exemplified by figures 2.3.4 – 2.3.7.

Other than establishing basic stylistic details about the graphical descriptors favoured by listeners, the examination of individual responses highlighted how listeners' depictions were modified according to their listening location. This occurrence (illustrated throughout appendix 2.B) was exemplified in figures 2.3.8 - 2.3.10, which depicted one listener’s responses to programme material reproduced using the ‘processed’ system setting. It was clear from these figures that the listener described the ensemble towards the front left of the vehicle when sat in the driver’s seat (figure 2.3.8), and towards the centre and rear of the vehicle when listening from the rear-centre and rear-left seats (figure 2.3.9 and 2.3.10).

It could therefore be suggested that descriptions were being influenced by listening location for this (and many other) listeners. Another determination from the examination of listeners' individual representations was that consistency was demonstrated in their graphical descriptions of the same stimulus. Although a lack of consistency did not necessarily reflect a failing of the listener or the descriptive process, if consistency was visible in a listener’s responses, the suggestion was that experiences were being reliably represented using the individual graphical languages. Consistency was illustrated by many of the depictions in appendix 2.B, and exemplified by listener 3’s descriptions from the rear-left listening location in figures 2.3.11 and 2.3.12.
Although individual listeners appeared consistent when representing their own experiences, differences were visible between listeners' graphical responses. One distinction was that certain listeners tended to position the cello (C) at the rear of the vehicle, as exemplified by figures 2.3.5, 2.3.9, 2.3.11 and 2.3.12 above. The cello was primarily described in this manner when the processed audio system was used; a possible reflection of more information being reproduced over the rear loudspeakers in this setting. Nevertheless, descriptions could not always be classified according to the audio setting used. For example, regardless of the system setting in use, listener 2 (exemplified in figure 2.3.5 and appendix 2.B figures 2.B.7-2.B.12) chose to dissect most ensembles and represent copies of the same instrument at various positions around the vehicle, typically where loudspeakers were located. Another notable contrast was in the size of listeners' descriptions. Experiences were represented using concise descriptors (for example figures 2.3.4 and 2.3.12) by some listeners whilst others covered larger areas of the vehicle with their depictions (for instance figure 2.3.9).

Thus it appeared that the reproduction system, listening location and even the listener themselves could have had an influence on the descriptions that were being elicited. It was therefore of interest to see what would be confirmed by a statistical analysis.

Statistical analysis
The statistical analysis was completed to identify whether listeners were describing differences in their experiences of ensemble width and location. In order to conduct this analysis, the selected attributes of listeners' responses were first converted into numerical data. For ensemble width, this meant finding the left-most point on each listener's depiction and measuring from this position to the right boundary of the ensemble to arrive at a width in millimetres. Measuring ensemble location was more complex and involved the creation of a transparent grid divided into millimetres. The vehicle's outline, taken from a blank response sheet, was represented on the grid along with a line bisecting the car down the centre from front to rear. By placing the transparent grid over a listener's response, the centre of the depiction relative to the central reference line on the transparency could be measured. A measurement of 0mm indicated a response centred on the middle of the car, with negative values highlighting a left bias to the ensemble, and positive values denoting a right-of-centre skew. The larger the value in either direction, the greater the offset from centre. Due to the programme material being recorded with instruments positioned further to the left than right, there was automatically a lack of symmetry in the ensembles which resulted in a left bias being noted by the statistical analysis.

The next problem for the statistical analysis was identifying which factors to include in the evaluation of listeners' spatial descriptions. Since both the listening location and reproduction system setting were believed to be of influence when listeners depicted their experiences, both was included as factors in each analysis22. Programme material was not included as a factor, for although the positions of individual instruments within each ensemble had been varied to prevent listeners relying on their existing knowledge of how instruments should be located, the width and location of the actual

22 Descriptions of ensemble location and ensemble width were evaluated separately.
ensembles were not manipulated, and therefore differences were not anticipated in how listeners would describe the spatial characteristics of the various ensembles. The final factor to be included in both analyses was the listener, as the examination of listeners' depictions had ascertained that differences were occurring in how individuals described their experiences.

Accordingly, numerical measures of listeners' graphical descriptions were analysed to establish if any differences in ensemble width and location could be attributed to the various factors (listening location, reproduction system setting and listener). Although some of the numerical data violated the assumption of normal distribution, the evaluation employed a parametric analysis of variance (ANOVA) because:

(i) ANOVA was known to be robust to minor violations of assumptions (Howell, 2002, p340),
(ii) the number of responses within each level (for example each seat) of each factor (e.g. listening location) were equal and
(iii) there were at least 20 degrees of freedom for error in the univariate ANOVA.

Statistical figures and tables referred to in the following section can be found in appendix 2.C. When ensemble width measurements were analysed using an ANOVA (table 2.C.1 in appendix 2.C), significant differences were identified between listeners' individual depictions ($F(11,144) = 33.571$, $p<0.001$), between descriptions when listeners sat in different listening locations ($F(2,144) = 6.836$, $p<0.01$) and for an interaction between the audio system and listening location ($F(2,144) = 10.954$, $p<0.001$).

For ensemble location, a greater number of factors and factor interactions were statistically significant. Amongst these influential factors were all individual elements (listener, listening location and reproduction system setting), the interaction of listening location and system setting ($F(2,144) = 24.279$, $p<0.001$) and that of the listener and listening location ($F(22, 144) = 2.470$, $p<0.01$). This final interaction suggested that not only did listeners collectively describe the location of the programme material differently when sat in different listening locations, but that individual listeners chose to describe the ensembles differently from the various seats within the vehicle. Table 2.C.2 in appendix 2.C contains further details of the ANOVA when ensemble location was considered.

The individual factors of listening location, listener and system setting were identified as having a similar influence on ensemble location when considered with respect to the total amount of variance in the analysis model (table 2.C.2, values for Partial $\eta^2$)24. However, differences between listeners had a greater influence on the description of ensemble width than where listeners were sat, or the reproduction setting used (table 2.C.1). The comparative strength of the listeners' influence, over that of listening location or system setting, further supported the earlier suspicion that listeners were describing their experiences differently from one another.

23 According to Tabachnick and Fidell (1996, p71), if there are at least 20 degrees of freedom for error in a univariate ANOVA the $F$ test is said to be reliable to violations of normality of variables.

24 $\eta^2$ represents the squared correlation between the independent variables and any (dependent) variable influenced by the controlled manipulation of these variables (Howell, 2002). The closer $\eta^2$ is to ‘1’, the greater the correlation between dependent and independent variables.
An additional indication of differences between listeners’ descriptions, and a visual example of how ensemble location differed for each of the reproduction settings, was provided when responses from all listeners (in all listening locations) were overlaid in exemplary plots. In figure 2.3.13, all listeners’ descriptions of the percussion (from programme item 3) were plotted together when this source was reproduced over the stereo system (figure 2.3.13:1) or the processed system setting (2.3.13:2). When the processed setting was employed, the majority of listeners (illustrated by the slightly darker shading on each figure) described the percussion at the centre of the vehicle. This implied that the processing may have succeeded in shifting the instrument away from the loudspeakers for some listeners; a proposal substantiated by the clustering of responses at the front left and front right of the vehicle in figure 2.3.13:1 when the stereo system was used. When depictions of the cello from the same stimulus were plotted together in figure 2.3.14, a greater spread of responses was visible across the vehicle and less clustering existed at specific locations - particularly when the cello was reproduced using the processed setting (figure 2.3.14:2). When individual depictions were analysed, some listeners appeared to have described differences in the size of this instrument and the percussion within the ensemble. However, listeners had informally mentioned that they had found it more difficult to localise the cello than the percussion. Consequently, the individuality in responses may have reflected difficulties when listeners described this lower frequency more sustained instrument, a difference between the instruments with respect to their perceived size, or a combination of both factors.

Further to the individual influences of the listener, it was originally ascertained in the ANOVA that listening location affected descriptions of ensemble width, \( F(2,144) = 6.836, \ p<0.01 \). Yet a more detailed comparison of pairs of listening locations could not establish any significant differences in width when this listening location factor was altered. A brief revision of listeners’ individual

---

25 Although every effort was made when constructing these plots to represent listeners’ original descriptions, it should be noted that the cluster density of each plot was determined from listeners’ combined plots by the researcher.

26 Post hoc comparisons using the Games-Howell procedure (variance for ensemble width was not equal across all groups) did not identify any significant difference in the width of ensembles depicted from different listening locations.
A summary of three initial investigations

Chapter 2 - Development

descriptions confirmed that for many listeners, listening location made little difference to the width of their descriptions.

Although the analysis of individual factors provided an immediate indication of how listeners were describing their experiences, this analysis did not take into account how these factors interacted to influence listeners' descriptions. Since the interaction of listening location and reproduction system had already been identified as significant in the description of both ensemble width and location, it was useful to look into this in more detail.

Figure 2.3.15 Means and 95% confidence intervals for ensemble width depicted from different listening locations over different reproduction system settings

A graph of means and 95% confidence intervals for ensemble width as described from different listening locations and using the two system settings (see figure 2.3.15) determined that the interaction between rear-central seat and the stereo setting resulted in listeners widest descriptions. Nevertheless, when listeners moved off-centre to the driver's seat, their descriptions were significantly narrower using the same reproduction system, whilst ensembles reproduced using the system's processed setting were described with similar width regardless of listening location. This implied that listeners' descriptions were affected to a greater degree when the stereo reproduction setting was used as the processed system was able to promote similar experiences in listeners regardless of their location in the vehicle. Thus, the processed system appeared to have achieved its objectives for these listeners, with the provision of a comparably wide image from each seat.

An ensemble's location with relation to the centre of the vehicle was also influenced by the interaction of listening location and system setting. As with ensemble width, the means and 95% confidence intervals for this interaction were plotted (see figure 2.3.16).

---

27 In figures 2.3.15 and 2.3.16, ensemble width (figure 2.3.15) and ensemble location (figure 2.3.16) are displayed as percentage values. In figure 2.3.15 100% refers to the width of the entire vehicle, whereas in figure 2.3.16 100% denotes an ensemble which was depicted on the right boundary of the vehicle. In the same figure, 0% indicates that an ensemble was depicted at the centre of the vehicle and '-100%' defines an ensemble depicted with a fully left location.

N Ford Doctoral Thesis 82
A summary of three initial investigations

Chapter 2 - Development

Figure 2.3.16 Means and 95% confidence intervals for ensemble location depicted from different listening locations over different reproduction system settings

![Graph showing means and 95% confidence intervals for ensemble location from different listening locations and system settings.]

An examination of figure 2.3.16 identified that listeners described a significantly greater skew to the left from the rear-left or the driver's seat and auditioning material using the system in its stereo setting, than when the processed system was used in combination with these sub-optimal listening locations.

This effect was also demonstrated in listeners' individual depictions, for example, figures 2.3.17-2.3.22 below. In these figures, listeners' responses from the off-centre driver's seat illustrate how the auditory scene skewed further to the left when the stereo system setting was used than when the processed setting was employed. However, even though processing succeeded in reducing the amount of skew described by listeners, the difference in descriptions from the central listening location and the off-centre driving location was still significant.

Figures 2.3.17 - 2.3.22 Descriptive differences when system setting changed from stereo (s) to processed (p) mode
Figure 2.3.17 L3s Figure 2.3.18 L3p Figure 2.3.19 L2s Figure 2.3.20 L2p Figure 2.3.21 L5s Figure 2.3.22 L5p

Since material reproduced using the processed setting was described as further away from the loudspeakers and of a constant width when participants were sat in all listening locations, the implication was that multichannel processing had reduced the influence of the precedence effect for listeners involved in this study and succeeded in enhancing the spatial characteristics being experienced.
2.3.3 Summary of findings
The principal objective of this investigation was to address the limitations identified by the pilot study. Developments were required with relation to the number of listeners involved, the number of repeated presentations of the same stimulus and the choice of programme material. Accordingly, this first formal investigation used an increased number of listeners, from three to 12, and obtained representations of selected spatial attributes\(^{28}\) from these listeners. To address the ambiguities which appeared to have been caused by the programme material in the pilot investigation, novel (simpler) stimuli were created. The use of trio ensembles resulted in listeners describing the same number of instalments within each depiction (it was further noted that this more extensive group of listeners used predominantly similar styles of representation to those employed by the original listeners), making the analysis of listeners' responses easier for the researcher. The final development prompted by the pilot investigation was the inclusion of a greater number of repeat runs. The analysis of listeners' repeated descriptions established that, where experiences were stable, these could be consistently described using individual graphical languages (I-GAL's).

The emphasis of the remaining analysis was on confirming whether listeners' graphical descriptions reflected the differences in their listening location and the audio reproduction system. Descriptions were evaluated in their original graphical form and measured to obtain data which could be statistically analysed. Amongst other findings, it was determined that the interaction of listening location and audio system setting resulted in listeners depicting spatial characteristics differently. Specifically, when sat in sub-optimal listening locations, listeners' descriptions were less skewed when material was reproduced over the audio system in its multichannel processed setting than in stereo. Moreover, although listeners described ensembles as widest from the central listening location when the stereo setting was used, a move to the sub-optimal listening locations resulted in a narrowing of their depictions which was not observed when material was reproduced using surround processing. The visible and significant difference in listeners' descriptions of ensemble width and location led to the conclusion that the manipulation of the audio system and the listening location had influenced listeners' auditory spatial experiences. Consequently, it appeared that the processed setting resulted in more stable descriptions from off-centre listening locations than the stereo setting.

Although participants were collectively describing differences when listening from different locations to material reproduced using the two system settings, the analysis of individual listener descriptions highlighted significant differences between listeners. It was noted that listeners, although using a predominantly similar depiction style, did use differently sized descriptors. A further examination of listeners' responses identified that some of the differences may have been due to the cello instrument within the trio ensembles. The increased spread when selected listeners' described this instrument originally suggested that the cello was being depicted larger to reflect their experiences. However,

\(^{28}\) The decision to describe the specific spatial attributes of ensemble width and ensemble location was taken as a means of focusing listeners' attention. It is once again acknowledged that this decision may have prevented listeners' from describing more pertinent aspects of their experiences.
when the spread of the cello descriptions was coupled with some listeners' informal comments, it suggested an alternative reason for the differences in listeners' descriptions; that the low frequency sustained nature of the source was problematic for listeners when attempting to describe their experiences. Whatever the reason, this ambiguity made it difficult for the researcher to have confidence in the comprehension of listeners' auditory spatial experiences through their graphical descriptions and confirmed the need for further inquiry.

In addition to addressing the potential for attributes of the programme material to cause ambiguities in the researcher's understanding of listeners' experiences, in the continued development of the descriptive graphical language, it was advisable to consider the limitations of the current listening population. Although more listeners were used in this investigation than the pilot study, it was still possible to censure this study for its use of a relatively small group of mainly experienced participants. Even though these more experienced listeners had demonstrated consistency when describing their experiences, it could not be known whether these descriptions would be representative of those obtainable from a more diverse, less expert, population and consequently, the external validity of the language was limited. It was also noted that ten of the 12 participants had played an active part in developing the automotive audio system used throughout the investigation, and although listeners were not familiar with the purpose or conditions of the study, their knowledge of the reproduction system could have informed their description. It was therefore concluded that any further investigation would attempt to establish how less experienced listeners (both in terms of their knowledge of critical listening and their awareness of the reproduction system) would describe their experiences using their own individual graphical descriptors.
2.4 The further development of a descriptive graphical language (GAL)

The third investigation in the continued development of a graphical language (GAL) sought to address the limitations and ambiguous elements from earlier studies. The listening population was increased to include 31 participants and the programme material was further simplified to identify how listeners were representing the lower frequency cello instrument compared with more transient sources. Although changes were made, several similarities remained between this and the earlier formal investigation. One similarity was in the use of a stationary vehicle equipped with the same audio system. However, differences between the reproduction settings were less obvious in this investigation. As a result, it was appropriate to ascertain whether listeners would still reflect these physical differences in their graphical descriptions. As with previous investigations, an examination of listeners' individual depictions was conducted alongside a basic statistical analysis. Further details of the investigation setting, procedure and the analysis of listeners' graphical descriptions are outlined in the following sections.

2.4.1 Details of the investigation setting

Audio reproduction system

Although this investigation took place within a different vehicle, the same multichannel surround processing audio system existed. Rather than use this system in its stereo and processed settings (the previous investigation had already indicated that listeners were reflecting differences between these settings in their graphical descriptions), it was of interest to verify whether listeners would describe differences in their experiences when differences between the system's settings were less obvious. Consequently, it was decided that the system would only be used in its processed setting, which up-mixed the two-channel information from a conventional CD into the different loudspeakers within the vehicle to obtain enhanced spatial information. This processed setting was further manipulated to obtain three distinct listening conditions: The first condition employed the centre loudspeaker with a full frequency range (the optimal setting for the audio system); the second restricted the signal that was reproduced by this loudspeaker so that only information above 1kHz was replayed; and the third removed this central loudspeaker from the system and relied on the generation of a phantom centre image from the front left and right loudspeakers.

Programme material and the creation of stimuli CDs

When trio ensembles had been reproduced in the previous investigation, the transient percussive instrument appeared to have been depicted within a smaller area of the vehicle than the lower-frequency continuous cello source by some listeners. It was therefore necessary to establish whether difficulties when localising the cello had caused the instrument to be described with greater variability, or whether listeners were simply describing the cello as less focused. The specific aim in determining the reason for the differences in description was to identify whether the choice of programme material could have caused listeners to represent their experiences differently from one another; a significant

29 The audio system's processed setting was described earlier in section 2.3.1 and by figures 2.3.1 and 2.3.2.
A summary of three initial investigations

Chapter 2 - Development

source of variability in the previous investigation. Clarification was required because any ambiguity could prevent the researcher from understanding listeners' experiences via their graphical depictions.

Four solo instruments (including voices) were used to create the two-channel stereo programme material employed throughout the investigation. As with the earlier automotive investigation, each anechoic mono instrumental recording was chosen from the *Archimedes* CD[^30]. The extracts selected from this CD once again included the sustained cello melody (with its frequency spectrum predominately in the low frequency region) and the transient higher frequency percussion. But, rather than combine instruments into more complex stereo scenes, each instrument was amplitude panned to occupy a central location in its own two-channel stereo recording. The decision to place instruments at the centre of each stereo scene was taken with reference to the different audio system settings and how these could affect listeners' descriptions. Specifically, the precedence effect appeared a likely reason for listeners shifting their descriptions towards the door speakers in the previous investigation when the centre loudspeaker was absent and listeners positioned in off-centre listening locations. Consequently, it was of interest to determine the cause of this effect when the centre channel was manipulated and the programme material had been edited so that individual instruments were at the centre of each stereo scene.

Individual programme items were edited to between 30 and 40 seconds in duration (depending on the melodic line) and, as in the previous investigation, an equal amount of artificial reverberation was added to all extracts so that they more closely resembled typical two-channel stereo material. A table of programme material is provided in table 2.4.1 below.

<table>
<thead>
<tr>
<th>Programme</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female speech</td>
<td>Mid frequency female speech extract in Danish</td>
</tr>
<tr>
<td>Cello</td>
<td>Low frequency continuous cello passage</td>
</tr>
<tr>
<td>Percussion</td>
<td>Transient percussive extract consisting two drums</td>
</tr>
<tr>
<td>Male speech</td>
<td>Low frequency male speech extract in Danish</td>
</tr>
</tbody>
</table>

Six CDs of programme material were recorded, each with its own running order to minimise the possibility that listeners' responses would be influenced by the presentation sequence of the material. Specific tracks appeared twice on the different CDs so that the repeated presentation of these items could provide an indication of the listeners' consistency in using their descriptive graphical languages.

Participants

To increase the external validity of any conclusions which could be drawn from the study, 31 listeners were asked to participate. Within this extended population, 12 listeners had been involved in the earlier descriptive study and therefore had (marginally) more experience of graphically describing their experiences than the 19 new listeners. Because a listener's previous knowledge of listening and

[^30]: See Hansen and Munch (1991) for details about the recording.
familiarity with the automotive audio system could have affected their graphical descriptions, all participants were asked to respond to a set of questions\(^3\) which identified their relevant histories. Answers to the questions revealed that, of the 12 listeners who had participated in the earlier investigation, nine had experience as musicians or listeners\(^3\) and only one had a job which did not involve a practical understanding of the audio system or its component parts. Conversely, only six of the new participants played an instrument or had previous listening experience. Moreover, with the exception of two listeners, new participants worked as designers or administrators and had little hands-on involvement with the design or manufacture of the audio system. When compared with the listening population from earlier investigations, this expanded group of participants could be considered a more varied and less experienced population. Therefore, responses from these listeners would provide an indication of whether a descriptive graphical language could be employed by a more diverse listening population.

**Investigation procedure**

Prior to their participation in the investigation, and regardless of their familiarity with graphical description, all listeners were given the same instruction sheet (see appendix 2.D). The information sheet reminded listeners that no correct responses existed since it was a description of their experiences that was important. Listeners were also informed of the investigation process and the spatial attributes to focus their attention on; namely the width and location of the individual instrument stimuli. Since listeners had previously described differences in the spatial attributes of ensemble width and location\(^3\), it was believed that by asking listeners to describe instrument width and location, they would have appropriate spatial attributes to concentrate on. Further reasons for the selection of these spatial attributes existed: Firstly, by obtaining descriptions of both characteristics from listeners it would help clarify the ambiguity from the earlier investigation; namely whether difficulties localising the cello in the trio ensemble had contributed to the variability between listeners' responses, or whether the cello was perceived - and subsequently described - as wider than the more transient instruments. The second reason for selecting these spatial attributes was to ascertain whether the comparatively small differences in audio system setting (the use of an optimal centre channel, a reduced bandwidth centre, or the generation of a phantom centre) would have an influence on specific features of listeners' auditory spatial experiences.

In addition to varying the signal being fed to the centre loudspeaker, listeners were asked to move between two listening locations, namely the driver's seat (front left) and the front passenger's seat. Since both locations were off-centre, varying this factor was not expected to be influential with respect

\(^3\) Amongst other questions, listeners were asked if they play(ed) musical instruments, the requirements of their jobs and any previous involvement in listening tests (including any training they may have received).

\(^3\) Although it is acknowledged that the experience acquired via critical listening is distinct from that obtained through the practice of playing a musical instrument, an ability to listen is fundamental for the successful completion of either task and both have been found (Bech 1989) to be influential in the subjective evaluation of reproduced audio.

\(^3\) The single distinction between ensemble width / location and instrument width / location is the type of stimuli being described; an ensemble of instruments or a single source.
to the width of listeners’ descriptions. However, the interaction of listening location and reproduction system setting had been influential when listeners had previously described ensemble location. The reduction in physical differences between the various system settings therefore provided a further opportunity to establish whether this (listening location and reproduction system) interaction would be statistically significant.

Figure 2.4.1 Graphical response sheet

Each listener participated individually in the study. At the start of each run, listeners were presented with their blank (A3) response sheets which consisted of a scaled representation of the vehicle (figure 2.4.1). Other than to illustrate a different make of vehicle, the graphical response form was the same as previously employed. Moreover, listeners were once again asked to provide descriptions of their auditory spatial experiences on this two-dimensional sheet using their own descriptive graphical languages. Listeners were not prescribed a graphical depiction style, rather they were told to use any graphical descriptors they wished so long as this enabled them to describe their experiences. During each run, a single item of programme material from one of the six CDs was repeatedly reproduced using the same audio system setting. To complete a single run, each listener was required to listen to this one extract and provide depictions on two response sheets; one description per listening location. After completing descriptions from both driver’s and passenger’s seats, listeners were asked to leave the vehicle to give the researcher time to change the CD track and reset the audio system (controlled using a laptop computer) ready for the next run. Listeners were informed they could take a break whenever tired, however most listeners completed the investigation within one hour and so declined this offer.

2.4.2 Analysing listeners’ responses

A combined statistical analysis and examination of individual listeners’ graphical descriptions provided the means of evaluating listeners’ responses. The principal objectives of this combined analysis were to identify whether:

(i) listeners were describing their auditory experience of the cello extract wider than the more transient programme items;
(ii) there was any indication of listeners being unable to localise the more sustained cello source;
(iii) the choice of programme material had any other influence on listeners’ responses;
(iv) listeners reflected any of the (comparatively small) differences between audio system settings in their graphical descriptions;
(v) individual listeners demonstrated consistency in their descriptions;
(vi) any ambiguities existed in listeners’ responses which would prevent the researcher from understanding the listeners’ experiences through their individual graphical descriptions;
Analysis of individual responses

The examination of individual listeners' descriptions concentrated specifically on three of the requirements of the analysis, namely (i) whether sources of ambiguity existed which could prevent the researcher from understanding listeners' experiences, (ii) whether there was any indication of listeners being unable to localise the more sustained cello source and (iii) if listeners were consistent in their graphical descriptions. An initial examination of listeners' individual descriptions established that, as with the previous investigation, differences existed between listeners' responses. For example, when listeners described their experiences of the same instrument (for instance the percussion using the unrestricted optimal centre channel in figures 2.4.2 and 2.4.3) visible differences existed in the size of the graphical descriptors being used. This finding led to the initial suggestion that listeners were employing different scales when representing their experiences on the graphical response form, or that their experiences were indeed different. A separate inquiry established that a few listeners were once again representing copies of the same instrument at various positions around the vehicle, and (as exemplified in figure 2.4.4) descriptors were typically positioned where the loudspeakers were located. It was acknowledged by the researcher that more clarification would be required to identify reasons for the variable size of descriptors and the division in the description of individual instruments.

As illustrated by figures 2.4.5 and 2.4.6, descriptors also varied within a listeners' descriptions whilst their individual graphical languages were being developed, tending to simplify over time. The development of descriptors over the duration of the investigation did not necessarily prove problematic in the understanding of listeners' auditory spatial experiences because listeners tended to maintain some consistency when portraying the location and width of an instrument. An exception to this statement was provided by a single listener who was unable to settle on a single graphical representation style (see appendix 2.E figures 2.E.38-2.E.53). This particular listener also appeared to describe little difference in the width of any stimulus regardless of the listening condition. Moreover,

---

34 To accompany this analysis, exemplary graphical descriptions from individual listeners are included in appendix 2.E. However selected descriptions are once again provided within the text for immediate reference.
additional descriptors in both this\textsuperscript{35} and a further listener's\textsuperscript{36} descriptions suggested that something other than their experience of instrument width or location may have been prompting their responses (see figures 2.4.7 and 2.4.8). Further ambiguities were present in the responses of other listeners. Whereas the majority of participants described the various solo instrument stimuli at the front and occupying the same side of the vehicle as where they were sat, the examination of individual descriptions identified that a small group of listeners were drawing instruments behind them or on opposite sides of the vehicle. Examples of this occurrence are provided in figures 2.4.9 - 2.4.11 below and in appendix 2.E, figures 2.E.36 - 2.E.47.

Where listeners' graphical representations deviated from those of the listening majority, it was still possible that these listeners were describing their experiences. However, it was also possible that some other factor could have led to their ambiguous responses. Whatever the reason for the differences in description, ambiguities in listeners' graphical representations were cause for concern, as the researcher's understanding of the listeners' experiences could be jeopardised when graphical representations were anomalous.

Although differences existed between listeners, an examination of individual responses suggested that, other than in extreme cases, listeners were able to demonstrate an element of consistency in their representations (see for example figures 2.4.12 and 2.4.13). Furthermore, as exemplified throughout appendix 2.E, inconsistencies that did exist were not specific to the cello; listeners did not tend to systematically describe the cello source any differently to the remaining instruments.

\textsuperscript{35} As illustrated in figure 2.4.7 and appendix 2.E (figures 2.E.48 - 2.E.53) listener 1 included large and small circles, arrows and letters in his descriptions.

\textsuperscript{36} Listener 17 used text on his response sheet. This text not only indicated the width of the sound, but a 'sound input' and the 'direction of sound' (see figure 2.4.8).
Essentially, consistent listeners maintained their descriptive consistency throughout the investigation and, where ambiguous responses were elicited, these were elicited regardless of instrument.

Statistical analysis
To provide the means of conducting the statistical analysis, depictions from each listener were measured, with all measurements made in millimetres. Instrument width was measured as the distance between the furthest left of a listener's depiction and its right most boundary. An instrument's location was measured with respect to its displacement from a designated reference position; a line down the centre of the vehicle running from front to back. To obtain this measure, a grid (in millimetres) was placed over each listener's description of an instrument, with the central reference position on the grid (0mm) aligned with the centre of the response sheet. A measure was then taken of how far the centre of each instrument description was displaced away from this central reference. Accordingly, instrument location was either a negative or positive value, depending on whether the listener had skewed their description to the left or right of centre.

The first requirement of the statistical analysis was to identify whether listeners were describing their experiences of the sustained cello stimulus wider than the more transient instruments. A second query requiring clarification was whether the programme material had any other significant influence on listeners' responses (whether, for example, programme material affected listeners' descriptions of instrument location) and a final concern was whether listeners were able to reflect the comparatively small differences in the audio system's settings in their graphical depictions. A parametric Analysis of Variance (ANOVA) was therefore conducted37 for all factors (programme material, listener and reproduction system38) using the width measurements taken from listeners' graphical descriptions. Results of this analysis indicated that, of these three factors, the listener caused the greatest variability in the description of instrument width, but all three factors were highly significant39. That the listener was largely responsible for the variability in the analysis was not surprising considering the differences and ambiguities that had already been identified when listeners' graphical descriptions were examined. The significance of the listener in the statistical analysis did, however, confirm the need for a formal inquiry into the differences between listeners' descriptions.

When a graph of means and 95% confidence intervals for the programme material factor (F(3,372) = 48.153, p<0.001) was plotted (see appendix 2.F, figure 2.F.1), it was clear that the cello was being described as significantly wider than the percussion and vocal material. However, this was something of a simplification. Results of the earlier investigation had ascertained that listeners' ensemble width

37 Although some of the numerical data violated the assumption of normal distribution, the evaluation employed a parametric analysis of variance (ANOVA) because of the robustness of the method to minor assumption violations (Howell, 2002, p.340). Moreover, the number of values within each level of each factor was equal and there were at least 20 degrees of freedom for error.
38 A second precautionary analysis was undertaken which confirmed the suspicion that the description of instrument width would not be influenced by listening location.
39 The ANOVA table is presented in appendix 2.F (table 2.F.1).
descriptions were being influenced by the audio system. Even though differences between the system’s settings were less obvious in this current investigation, the system was still a potential influence on listeners’ descriptions of instrument width. The analysis of variance confirmed that not only was this system factor significant \( F(2, 372) = 59.019, p<0.001 \)\(^{40} \), but that the interaction of programme material and system setting was also found to be significant \( F(6, 372) = 5.617, p<0.001 \). Accordingly, a graph of means and 95% confidence intervals for the system by programme interaction was plotted and is included in figure 2.4.14.

Figure 2.4.14 Means and 95% confidence intervals for instrument width depicted for different programme items over different reproduction system settings

From this graph it was determined that only when the reduced bandwidth centre channel was employed, or a phantom centre generated, was the width of the cello instrument described as significantly wider than the remaining instruments. An inspection of the graph also highlighted that the width of the cello was not being influenced by the system setting, but that the more transient instruments were being affected; instrument width described as significantly narrower when the phantom centre, rather than the optimal setting was used. Even though the female voice and the percussion were described as significantly narrower than the cello in the reduced bandwidth setting, no difference in width was effected by the system when in either its optimal or reduced bandwidth setting.

The findings of the statistical analysis could also be illustrated in figures 2.4.15 - 2.4.18. Within these figures, responses for a selection of (the more typical) listeners were overlaid and plotted together, giving an indication of the density of responses at a particular location. These exemplary plots highlighted how cello and percussion descriptions from these listeners covered a similar area at the front of the vehicle when the optimal (full bandwidth centre) system setting was used. However, when the system was used with a phantom centre, the cello was described as covering a larger area than the percussion.

\(^{40} \) Instrument width was narrower when the audio system generated a phantom centre than when either a reduced bandwidth or optimum centre channel setting was employed.
Results of the statistical analysis suggested that, with the exception of the cello, the width of listeners' descriptions could be affected by the system setting, with the phantom centre resulting in a narrowing of individual instruments' descriptions. Since all instruments had been described with a similar width when the optimal system setting was employed, it was not possible to attribute all ambiguities from the earlier investigation (where descriptions for the cello had appeared to cover a larger percentage of the vehicle than the remaining instruments within the ensemble) to listeners describing the cello as significantly wider than the more transient sources. Therefore, the possibility still existed for variability in listeners' responses (F(30, 372) = 63.269, p<0.001) to be caused by something other than the cello occupying a larger space within the vehicle. The earlier examination of listeners' responses had indicated that, in the main, listeners were not describing the cello with any more variability than the remaining stimuli. Since the cello was not always depicted as wider than the remaining stimuli, and listeners' descriptions did not suggest any particular problems when localising this source, it was not possible to attribute differences in the previous investigation to the description of individual instruments within the programme material. Since the ambiguities in listeners' responses could not be directly attributed to programme material, it was acknowledged that further specific analysis would be needed to identify why these anomalies were occurring, as any ambiguity in representation could prevent the researcher from understanding a listener's experiences.

A second statistical analysis looked at listeners' descriptions of instrument location. The first factor to be included in the analysis was the listening position. By itself, this statistic was of little interest since both driver's and passenger's seats were sub-optimal and equal in terms of their proximity to the nearest loudspeaker and distance from the centre. However, the interaction of the sub-optimal listening locations with the reproduction system setting was of greater interest to determine whether listeners were able to describe the comparably small differences between system settings when sat in off-centre locations. Finally, programme material was included as a factor in the analysis. Although all material had been amplitude panned to the same central location within each recording, it was of interest to establish whether the sub-optimal listening locations would affect positional differences in how listeners described the various programme items. Furthermore, the inclusion of programme material enabled the influence of the listening locations on the skew of the different programme items to be evaluated with respect to the reproduction system.
The results of the ANOVA for instrument location are presented in appendix 2.F, table 2.F.2. Most variability within the ANOVA model could be attributed to listening location, and the factor was a highly significant cause of descriptive differences in instrument location \( (F(1,720)=1689.817, \ p<0.001) \). As previously mentioned, the analysis of instrument location with reference to a listener’s listening position was not particularly informative until this seating position was combined with the other factors within the analysis. The interaction of listening location with both programme material \( (F(3, 720) = 20.595, \ p<0.001) \) and reproduction system \( (F(2,720) = 71.929, \ p<0.001) \) was significant. However, the three way interaction of programme material, listening location and reproduction system setting did not significantly influence listeners’ descriptions. Graphs of means and 95% confidence intervals were plotted for both significant interactions.

Figure 2.F.2 in appendix 2.F illustrates how instrument location was depicted from different listening locations for different programme material. An examination of figure 2.F.2 (instrument location means and 95% confidence intervals for different listening locations and programme items) determined that the cello was being described significantly closer to the centre of the vehicle than the remaining instruments from the driver’s seat and closer to the centre (from all except the male voice) from the passenger’s seat. When combined with the earlier analysis of instrument width, it could be suggested that the increased width of the cello (when evaluated overall and not according to the individual reproduction system settings) was influencing the measurement of this instrument’s location.

The interaction of system setting and listening location (figure 2.4.19) confirmed that listeners were able to describe significant differences in the location of the programme material when this was reproduced using the different system settings. A review of figure 2.4.19 illustrated how listeners described significant differences between all reproduction system settings. Programme material was described as significantly closer to the centre of the vehicle when the optimal setting was used, whereas both reduced bandwidth and phantom centre settings resulted in listeners’ descriptions being skewed nearer to the side of the vehicle where they were sat and significant differences existed between these two sub-optimal reproduction system settings. Thus, even though the differences between system settings were smaller than in the

41 The influence of the listener was evaluated in a separate analysis of variance and found to be have a significant effect on the description of instrument location \( (F(30,496) =4.857, \ p=0.001) \).
previous investigation, listeners were still capable of reflecting these differences in their graphical descriptions.

2.4.3 Summary of findings

The third investigation in the continued development of a descriptive graphical language (GAL) once again sought to address the ambiguous elements noted in earlier studies. The listening population was increased and the programme material further simplified to identify how listeners were representing the lower frequency cello instrument compared with more transient sources. It was appropriate to ascertain whether the more diverse listening population would describe differences in their experiences when comparatively\(^42\) small changes were made to the audio system used to reproduce the simplified stimuli. As with previous investigations, an examination of listeners' individual depictions was conducted alongside a basic statistical analysis to provide answers to this and other queries. The combined analysis identified that:

(i) Listeners were only describing the cello wider than the more transient programme items when the sub-optimal audio system settings were employed. The listeners’ descriptions of the cello stimulus were not affected by the change in system setting (experiences of this instrument were described with the same width) rather, the vocal and percussion extracts were described as narrower when the reduced bandwidth centre channel and phantom centre were employed. Further to recognising that the cello was not uniformly described as wider than the other instruments, this analysis confirmed that firstly, listeners were able to describe differences in the width of instruments when relatively small alterations were made to the reproduction system, and secondly, that the full bandwidth centre channel setting appeared to produce the most stable graphical descriptions, as anticipated by Nind (2001).\(^43\)

(ii) The examination of graphical responses was not able to provide any evidence of listeners systematically describing the cello with more variance than other instruments as a result of their inability to localise this stimulus. Therefore it could not be confirmed that programme material was specifically responsible for ambiguities in listeners' descriptions.

(iii) When combined with listening location, the type of stimulus affected listeners' descriptions of instrument location. The increased width of the cello (when evaluated overall and not according to the individual reproduction system settings) appeared to be influencing the measurement of this instrument’s location.

(iv) In addition to describing width differently when the programme material was reproduced using the various settings of the reproduction system, listeners also described the location of their auditory experiences differently according to the system setting employed. The precedence effect (detailed earlier in this chapter) appeared a likely cause of listeners describing instruments closer to their nearest loudspeaker when the reduced bandwidth and

\(^42\) Small when compared with the alterations made to the reproduction system in the earlier investigation

\(^43\) Section 2.3.1 contains details of the multichannel audio system’s processed setting, and the supposed benefits of using this system within vehicles
phantom centre settings were employed. When the optimal reproduction system was used, listeners described the programme material towards the centre of the vehicle. Once again, this finding was anticipated by Nind (2001) and confirmed that (even though it was not the main aim of this thesis) listeners’ experiences, when represented graphically, could provide useful information for audio manufacturers.

(v) Other than when they were developing the descriptive languages, many listeners were consistent in the graphical descriptions. Although only a brief examination of listeners’ graphical responses was made with respect to individual listener consistency, this finding suggested that when a listener’s experiences remained stable, a descriptive graphical language could be used reliably.

Many positive discoveries emerged from this third investigation; one of these being that a more diverse listening population appeared able to use their own individual graphical languages to describe differences in their spatial experience. However, to recapitulate the research objective: the aim of the descriptive language was to enable the structuring and representation of listeners’ auditory spatial experiences in order that these experiences might be understood by the researcher. Thus, the existence of (unexplainable) ambiguities in listeners’ descriptions was considered the most important finding from this study, since these anomalies could prevent the researcher from understanding the listeners’ experiences through their individual graphical descriptions.
2.5 Conclusions

The early development of the descriptive graphical language (GAL) sought to obtain an elementary understanding of the spatial characteristics of listeners’ auditory experiences that could be structured and represented graphically. To achieve this aim, the three investigations shared a common emphasis in asking listeners to describe their own experiences using their own graphical terminology. In their individual graphical languages, listeners had a meaningful way of exploring and structuring their own experiences, and the descriptions elicited from individual listeners provided the researcher with inaugural evidence that auditory spatial experiences could be structured and represented graphically. Further to eliciting rudimentary graphical descriptions from individual listeners, each investigation sought to address limitations revealed by the previous graphical language studies, and in this way each continued to develop the investigation procedure, improving on the graphical descriptions that could be obtained from listeners and the means of analysing these responses. A brief summary of each investigation is provided below.

2.5.1 Summary of three initial investigations

The pilot investigation

The small scale pilot investigation provided the first indication of the auditory spatial experiences that could be graphically represented by (an albeit restricted group of) listeners. The investigation required that the three listeners graphically describe their experiences of complex two-channel stereo programme material when the loudspeaker location and their listening location were altered. By manipulating these variables, the pilot investigation was able to identify whether physical differences would influence listeners’ graphical descriptions. Listeners were not provided with a definition of what constituted a spatial characteristic, although (due to uncertainty about how to draw their experiences) listeners were (reluctantly) provided with a sketch of a possible drawing technique.

Depictions were analysed by the researcher to obtain a preliminary understanding of the spatial characteristics of listeners’ experiences that could be represented graphically. Three methods of analysis were employed: (i) an examination of listeners’ individual responses (to determine trends in depiction); (ii) the creation of overlaid plots of all listeners descriptions (to ascertain whether listeners were describing their experiences similarly) and; (iii) selected attributes (based on the conclusions of the graphical analyses) were measured to provide data for a rudimentary numerical analysis. It was confirmed that listeners were able to describe differences in the location (skew) and width of complex musical ensembles from various listening locations with the material reproduced from different loudspeaker locations. However, differences in listeners’ descriptions of other spatial characteristics (in particular ensemble depth) could not be explained by the manipulation of variables in this study. It was implied that the lack of coherent depth description (although a possible reflection of listeners’ individual experiences) could have resulted from listeners lacking a visual cue as to the location of the rear wall in the listening room. It was therefore determined that the response sheet would be improved by providing a scale model of the listening environment, thus aiding in the translation between listeners’ three-dimensional egocentric experiences and the two-dimensional plan representation.
An obvious problem at this early stage in the development of GAL was the limited listening population. As only two listeners were able to complete the full investigation (and both of these were experienced musicians) the results of this pilot study had severely limited external validity; it could not be known whether the auditory spatial experiences described by these listeners would be representative of a wider listening population. A second consideration was the complexity of the programme material as, other than for the simplest of material, listeners were not describing the same instruments as one another. Furthermore, the programme material in this study was commercially available, with the possibility that listeners’ descriptions were representing their pre-conditioned, rather than current, experiences. Thus, the simplification of the programme material and the use of created rather than existing material was advocated.

The final cause for concern identified by the pilot study was the investigation procedure itself. The investigation used only a limited number of runs, which only enabled listeners to describe one musical extract twice. Even though a listener was not expected to have the same experience each time they heard the same physical stimulus, the presence of consistency in listeners’ responses would have suggested a degree of reliability in the graphical medium for representing stable experiences. Consequently, it was considered important for subsequent investigations to contain a greater number of repeated runs.

Even though results suggested that listeners were able to describe some auditory spatial experiences using a graphical language, the elementary nature of the study was limiting and the conclusion at the end of the pilot investigation was that further research was required to address numerous causes for concern.

A formal study of graphical description
The second investigation addressed the causes for concern identified by the pilot study. Simple trio ensembles were created as programme material for this investigation and, to increase the listening population, 12 (mainly experienced) listeners were involved. Moreover, the graphical response sheet was designed as a scaled representation of the listening environment and included more visual landmarks to improve the translation from listeners’ experiences to response. To provide a focus for listeners’ attention when describing their experiences (and to reduce the amount of graphical data that would be elicited) listeners were asked to describe their experiences of ensemble width and ensemble location, the two attributes for which differences in listeners’ descriptions had been identified in the pilot investigation.

The investigation was the first to be conducted in a stationary vehicle equipped with an audio system capable of reproducing two-channel material over both multichannel stereo and surround processed settings. With listeners sat in three different locations (only one of which was optimal for stereo
listening) it was believed that their descriptions might reflect the different modes of the audio reproduction system.

Listeners’ graphical descriptions were examined to identify trends in responses and a statistical analysis confirmed if any statistically significant differences existed between listeners’ descriptions. It was identified that the same number of instruments were being depicted by each listener - a possible consequence of the simplified programme material - and that listeners used predominantly the same style of representation (either circles or rectangles). An examination of depictions established that listeners were consistent in the use of their own languages, suggesting an element of reliability to the descriptive medium. The statistical analysis determined that the interaction of listening location and system setting was significant. When sat in sub-optimal locations, listeners depicted the processed material closer to the centre of the vehicle than the stereo reproductions. Furthermore, other than when listeners were sat centrally in the vehicle, depictions were narrower using the system’s stereo setting. The visible and significant differences in listeners’ descriptions of ensemble width and location led to the determination that the manipulation of both the audio system and the listening location had influenced listeners’ auditory spatial experiences. The processed setting therefore appeared to result in more stable, spatially enhanced descriptions from off-centre listening locations than the stereo reproduction.

The examination of listeners’ descriptions also indicated that listeners were using differently sized graphical descriptors when representing their experiences, with the cello instrument from within the trio ensemble being described as spread over a larger area than the percussion. Since it was not known whether this spread resulted from listeners describing a physically larger source or, as suggested by informal comments, if it was caused by difficulties when listeners attempted to localise the instrument, it was difficult for the researcher to be confident in their comprehension of listeners’ experiences, and the need for further inquiry was confirmed. The listening population was, once again, cause for concern. Here, the use of a relatively small group of mainly experienced listeners meant that however promising the results, they might be found unrepresentative of a more generic sample of listeners.

The further development of a descriptive graphical language (GAL)
The third investigation in the continued development of a graphical language once again sought to improve on limitations and ambiguous elements from earlier studies. To increase the external validity of any conclusions drawn from the study, the number of listeners was increased to 31, of whom 19 were new to graphical elicitation, 15 were experienced and 10 were knowledgeable about the audio system used in this and the previous study. Solo instruments were chosen as programme material to ascertain whether this choice would cause ambiguities in listeners’ descriptions. Although the vehicle used in this study was different from that used in the earlier investigation, the same audio system was used, but with less obvious variations between system settings. Since the conditions of this investigation could be considered a modification of the earlier study, it was appropriate to ascertain
whether listeners would still experience, and therefore describe differences in instrument width and location when comparatively small differences existed between audio system settings.

As with previous investigations, an examination of listeners’ individual depictions was conducted alongside a basic statistical analysis. The combined analysis identified that:

(i) When programme material was reproduced using the full bandwidth centre channel, all instruments were described with the same width. However, in the sub-optimal system settings, vocal and percussion extracts were described as occupying a narrower space in the vehicle than the cello.

(ii) There was no evidence of listeners describing the cello with more variance than other instruments. Accordingly, differences between listeners’ descriptions could not be attributed to the choice of programme material.

(iii) Instrument location was described differently by listeners according to the audio system setting and programme item; auditory scenes were depicted significantly closer to the centre of the vehicle as the system setting used became more optimal, and (regardless of system setting) the cello was described closer to the centre of the vehicle than the other instruments.

(iv) Many listeners were consistent in their graphical description, and although only a brief analysis of response consistency was conducted, this finding suggested that when a listener’s experience remained stable, a descriptive graphical language could be used reliably.

Although positive findings were obtained from this third investigation - specifically that listeners were representing differences using their individual languages when small differences existed in the external stimuli - the existence of (unexplainable) ambiguities in listeners’ descriptions was considered the most important discovery of the study as these anomalies prevented the researcher from understanding the listeners’ experiences through their individual graphical descriptions.

2.5.2 The necessity for further work

Results of the initial investigations had identified that listeners were describing differences in their experiences when external stimuli were manipulated. Improvements were also made to the investigation process, with programme material simplified throughout the studies and the listening population expanded to included a greater number of inexperienced listeners. Moreover the initial investigations had established that graphical responses could be examined, overlaid descriptor plots created and statistical analysis conducted using a numerical measure of these descriptions; all of which could provide the researcher with information about listeners’ graphical responses. Thus positive findings could be gleaned from the three initial investigations. However, the studies had not only been tasked with obtaining an elementary understanding of the auditory spatial experiences that could be structured and represented graphically, but also to identify any anomalies which might be occurring during the descriptive process.
The initial investigations had indeed identified ambiguities in responses, with listeners not always responding similarly when presented with the same stimulus and their depictions not necessarily concurring with the descriptions expected by the researcher in response to the task set out in the investigation requirements. With the principal emphasis of the research directed towards an understanding of the listeners’ auditory spatial experiences, there was always the possibility that observed ambiguities were perfectly valid; relating to the different experiences of the listeners. However, the accomplishment of the research objective required the development of a descriptive language which would provide the researcher with an opportunity for obtaining a greater understanding of listeners’ experiences when these were communicated graphically. Consequently, even when listeners used their own individual GAL to represent their own auditory spatial experiences, the achievement of the research objective was not guaranteed, as ambiguities could be introduced at other stages of the communication which could still prevent the researcher from understanding the listeners’ experiences.

Figure 2.5.1 describes this problem graphically. In GAL development model stage I, a listener’s experiences (Exp¹) are able to be represented by the use of a listener’s own individual representations (IR). However, currently unexplained ambiguity (?) is introduced into the descriptive process which prevents the listener’s representations from being understood by the researcher. And, with the introduction of ambiguity, the possibility exists for the researcher to have an alternative comprehension (CM), a misunderstanding, of the listener’s experiences.

Figure 2.5.1 GAL development model stage I: The initial GAL investigations

The researcher misunderstands (CM) the listener’s auditory spatial experiences because (even though the listener is able to describe their experiences (Exp¹) using their individual graphical representations - IR) ambiguities elsewhere in the descriptive process hinder the researcher’s understanding of these graphically represented experiences. (Research objective not fulfilled)

A key for the descriptive process model is provided in appendix 1
Accordingly, with the validity of the evaluation determined by the "adequacy of the researcher to understand and represent meanings" (Banister et al., 1994, p143), and in line with the questions posed at the start of this thesis, it was essential to identify what the ambiguities were, why these ambiguities were occurring and where in the communication process they were appearing, in order that such anomalies could be minimised.

As identified in chapter 1, an acknowledged method of obtaining this form of response clarification was to return any findings to investigation participants, since "An alert and observant actor in the setting is bound to know more than the researcher ever will about the realities under investigation...In that sense, local informants can act as judges, evaluating the major findings of a study" (Miles and Huberman, 1994, p275). Chapter 3 therefore provides details of an investigation which sought to clarify listeners' individual graphical responses by returning descriptions to their originators.
Chapter 2 has provided a detailed summary of three separate investigations conducted at the earliest stage in the development of the descriptive graphical language.

The aim of each investigation was to obtain preliminary information about the auditory spatial experiences that listeners could represent graphically. Furthermore, each study provided an opportunity for establishing methods of evaluating the elicited graphical data and all contributed to the further development of the graphical language by highlighting considerations for the investigation procedure.

The rudimentary numerical analysis of the graphical responses from the small scale pilot study (detailed in section 2.2) ascertained that listeners were describing differences in the width and location of complex two-channel stereo programme material. However the complexity of this material and the simplicity of the response sheet appeared to cause problems for listeners and both were identified as considerations in the continued development of a GAL. A further cause for concern was the listening population, which required expansion to improve on the study’s limited external validity. A final consideration was the inclusion of a greater number of repeated runs in any further GAL investigations.

The second investigation (section 2.3) addressed the causes for concern identified by the pilot study, simplifying the programme material, increasing the listening population and number of repeated runs and improving the graphical response sheet. The examination of depictions identified that listeners were consistent in the use of their own graphical languages whilst the statistical analysis determined that the interaction of listening location and system setting was significant; with the processed audio system setting resulting in more stable, spatially enhanced descriptions from off-centre listening locations than the stereo reproduction. However, differences were identified between responses which made the task of the researcher more complex when trying to understand listeners’ experiences through their graphical representations. These descriptive ambiguities along with the use of a restricted (mainly experienced) listening population, were the main considerations in the design of the final investigation.

The third investigation once again sought to improve on limitations from earlier studies. The number of listeners was increased and included a greater number of inexperienced listeners and the programme material was further simplified to ascertain whether it could have contributed to the differences between listeners’ descriptions. A further manipulation was the audio system, which was used with comparatively small differences between the reproduction settings. The analysis of listeners’ responses established that listeners were still describing differences in the width and location of the solo instrument stimuli when the audio system was manipulated. However, the existence of unexplainable ambiguities in listeners’ descriptions continued to be problematic as these could prevent the researcher from understanding the listeners’ experiences through their individual graphical descriptions.

A further investigation was proposed to clarify listeners’ individual graphical descriptions.
CHAPTER 3

Clarifying individual listeners' descriptions
The development of a universal graphical language
3.0 Chapter overview

Chapter 3 is divided into two parts. In the first part of the chapter (clarifying individual listeners' descriptions), a summary is provided of an investigation which sought to clarify the unexplainable ambiguities present in listeners' individual graphical descriptions at the end of the initial investigations. Briefly, the clarification study involved two phases; (i) eliciting graphical descriptions from individual listeners and (ii) returning these responses to their originators to obtain a verbal clarification of what was being depicted, in order to identify sources of ambiguities.

Prior to detailing the clarification investigation settings (in section 3.2) and pertinent observations from this study (in section 3.3), the introduction of chapter 3 (section 3.1) summarises accepted causes of anomalies in subjective sensory evaluation. The most relevant causes (for a listener who uses their own descriptive language to represent their own auditory spatial experiences) are then referred to in the subsequent account of the clarification investigation.

In section 3.3, details are provided of the verbal descriptions that were offered by listeners when asked to explain (during the clarification process) what they were representing in their graphical descriptions. Sections 3.3 and 3.4 (the summary of findings) explain why - when asked by the researcher to describe the width of reproduced ensembles and solo instrument stimuli - listeners may have been required to represent attributes not readily experienced or easily represented using a graphical medium.

The discoveries identified during the clarification investigation are illustrated in the descriptive process model at the end of section 3.4.

Although the use of appropriate written instruction is identified as a means of reducing anomalies in listeners' descriptions, section 3.5 presents a rationale for the continued development of the descriptive graphical language. Specifically, without a continual process of clarification, the researcher is always in a position where the misunderstanding of listeners' experiences is possible, if experiences are communicated using an individual language.

To resolve this problem, the development (by listeners) of a mutually acceptable language of graphical descriptors to represent their individual auditory spatial experiences is proposed in section 3.5. The process of developing this universal graphical language (U-GAL) is summarised in the second part of chapter 3 (sections 3.6 – 3.8).

Listeners were initially asked to participate in small discussion panels to develop both the U-GAL and a set of accompanying verbal descriptors; this verbal language was evolved to enable to provision of suitable written investigation instruction by the researcher. An overview of the language development process is provided in section 3.6, along with a précis of the individual panel discussion sessions.
Following the development of the individual panel languages, one listener from each panel was selected by their peers to discuss their panel’s graphical and verbal descriptors in the final stage of U-GAL’s development. The development process is summarised in section 3.7, and the universal language of graphical descriptors developed by these listeners is introduced in the same section.

Key elements and principal findings - from the clarification of listeners’ individual graphical descriptions to the development of the inter-subjective U-GAL - are outlined in the summary and conclusion of chapter 3 in section 3.8.

At the end of this final section, figure 3.8.1 illustrates the descriptive process model at this stage in the graphical language’s development.
3.1 Introduction

Conclusions of previous investigations established that many of the participating listeners were able to consistently describe their experiences of the selected spatial attributes using their own individual graphical descriptors. However, at this stage of the development process it is worth re-stating the research objective. Specifically:

*To develop a descriptive graphical language which enables the structuring and representation of listeners' auditory spatial experiences in order that these experiences may be understood when communicated to the researcher.*

Although many listeners demonstrated consistency when describing their experiences in the initial studies, ambiguities did exist in listeners' graphical descriptions. Through the analysis of these descriptions it was identified that the representations of a selection of listeners did not match those of the remaining investigation participants; nor did these concur with the descriptions expected by the researcher in response to the task set out in the investigation requirements. Any ambiguity in a listener’s graphical descriptions makes the task of the researcher more complex when attempting to understand the listeners' auditory spatial experiences through their representation. Consequently, with the existence of any ambiguity, the suitability of the graphical medium for fulfilling the objective of the research can still be disputed; it is therefore prudent to identify at this stage why ambiguities occur.

3.1.1 Identifying sources of ambiguity in graphical responses

Throughout the history of sensory analysis, researchers have been faced with the challenge of attempting to obtain descriptions of events (external stimuli or internal experiences) or statements of preference from individuals. However, individuals are precisely that - *individual*. And although there may be a communalisation of experiences within a group of individuals, ambiguities are more likely to occur when a response is sought from these mutable evaluators than from a calibrated measuring device. Using contemporary research as a reference, specifically Köster (2003) and Meilgaard et al. (1999), the following sources of response ambiguity are acknowledged:

1) Differences in the experience of the individual: When provided with the same external stimulus, an individual may respond differently to their peers or than expected.

2) Differences in deciphering the requirements of an investigation: Individuals may understand the requirements of a study differently from how these were intended to be understood. Thus listeners may solve a different problem to that actually proposed.

3) Differences in responding in the language of the investigation: Although they may know what problem they are required to solve, a listener may find the response method difficult to use.

\[1\] The spatial attributes requiring graphical description by listeners were ensemble width and location and individual instrument width and location
4) Differences in how the investigation is completed: A listener may understand the problem they are required to solve and may be able to use the response language; however, they may solve the problem in a different way to expected.

5) Differences over time: It is possible that an individual will change their way of responding over the course of an investigation and consequently produce responses which are inconsistent.

Even though it is simpler to establish the reliability of a graphical language when a listener responds consistently to the same stimulus, it is worth noting that in response to this fifth source of ambiguity, inconsistency is not necessarily an indicator of a listener’s inability to represent their experiences graphically. Indeed Köster suggests that looking for consistency in an individual’s responses is one of the great fallacies in sensory science, as each individual:

Has a memory and therefore the second encounter with a stimulus may not mean the same to him or her as the first one. (Köster, 2003, p361)

Thus, although a listener’s representations may at first appear inconsistent, if inconsistencies in description are a result of differences in a listener’s experiences over time - for instance, a difference in the way the same stimulus is experienced each time it is encountered - a listener’s responses can still reflect their experiences. However, should the change in description be motivated by some other factor, such as the investigation setting itself, differences become more of an issue since a listener’s descriptions may no longer represent their experience. For example, to combat boredom or tiredness (or as a result of their acquisition of knowledge and confidence) a listener may become more inventive in their use of language as an investigation progresses and consequently alter their description. Inconsistencies in description which occur as a result of the investigation setting can be reduced if listeners encounter stimuli in different orders and take breaks at regular intervals. But, as indicated by the five sources of ambiguity noted above, it is not just order effects which influence responses in sensory investigations. The first three sources of ambiguity are all cited in the work on Sensory Evaluation Techniques by Meilgaard et al. who declare:

The annals of sensory testing are replete with results that are unreliable because many of the panelists did not understand the questions and / or the terminology used in the test, did not recognize the flavor or texture parameters in the products, or did not feel comfortable with the mechanics of the test or the numerical expressions used. (Meilgaard et al., 1999, p2)

An individual’s inability to recognise attributes of a stimulus constitutes an obvious problem for descriptive analysis, when this stimulus - the product - is the external reference against which a panelist’s responses are validated. However, as Meilgaard once again asserts:

Through training and the use of references we can attempt to shape the mental process so that subjects move toward showing the same response to a given stimulus. (ibid. p3)

Shaping the mental process of an individual is advisable in situations where participants are required to evaluate an event which is external to themselves: Participants in descriptive studies can be taught how to respond to a particular event and in this way can learn to recognise a particular object and increase
their consistency as part of a panel. However the shaping process is at odds with the aims of a GAL. Put simply; training a listener to recognise an object influences the listener's experience of that object. Thus, following training, a listener’s representations would be more likely to reflect the investigation process and the experiences of the researcher, rather than their own experiences. When establishing the success of a GAL it is of little consequence whether a listener is able or not able to recognise spatial attributes of a stimulus - so long as their graphical descriptions reflect their actual experience and are not an outcome of participating in the investigation.

Potentials for ambiguities in communication have been covered extensively in Chapter 1. Be these from researcher-to-listener or listener-to-researcher, ambiguities and anomalies are problematic for sensory research where conclusions are based solely on the responses of the participant. Should the respondent not feel comfortable with the terminology used in the question asked by the researcher or the mechanics by which they are required to respond, there exists the potential that; (i) the individual will be unable to describe their actual experience, or that; (ii) the experience described by the listener does not answer the question posed by the researcher - a problem since the researcher will look to this question when providing a descriptive account of the listener’s representations. Ambiguities can be minimised if the investigation uses language familiar to the respondent. But what constitutes familiar? As Guski declares:

> It should be noted that individual human subjects in psychoacoustic laboratories still have their individual history and may use even common language in a slightly different way than the experimenter intends. (Guski, 1997, p767)

Kelly believes the terminology used by an individual - for instance a graphical description of an experience using their own representational style - will have meaning for them. However the meaning associated with this terminology may not be universally held and consequently the terminology may represent an alternative experience in another. He asserts:

> Since constructs are primarily personal, not all of them are easily shared. The particular nature of a person’s construct or his unusual use of terminology may be misleading to his listener. (Kelly, 1963, p116)

Since individual listeners have been using their own individual graphical languages to represent their own experiences, it is believed that little ambiguity should have been introduced during the structuring and representation of their experiences. However, the possibility still exists for the researcher to misunderstand the experiences being communicated by listeners and for the listener to misconstrue the requirements of the investigation.

In GAL investigations to date, written instructions have been provided for the listeners. Although brief, these instructions have included verbal descriptions of the spatial characteristics to be focused upon and visually depicted. However the language used in this verbal directive may have resulted in listeners each choosing a unique way to solve the investigation problem and depicting experiences based on their own interpretation of the investigation requirements.
Thus listeners may, or may not have a different experience when presented with the same stimulus. They may choose to represent this experience differently from one another using their own language and may even alter their descriptions mid-way through an investigation to reflect a change in their experiences. Although these are all valid reasons why differences may exist in a listener's descriptions, it is not yet known whether current ambiguities are a result of a valid difference in a listener's experience or a consequence of some unwanted variability; for example the listener failing to understand the requirements of an investigation. Since any unwanted ambiguity jeopardises the researcher's capacity to describe a listener's experience through their representations - and as a result threatens the realisation of the research objective - a further investigation was conducted to clarify where ambiguities existed in listeners' individual representations and why these were occurring.
3.2 Clarification process

A two-part study was therefore devised and completed to understand more about each listener’s graphical depictions. In the first part of the study, individual listeners were once again asked to describe spatial attributes experienced when items of programme material were reproduced within a motionless vehicle. As with previous studies, listeners were instructed to describe their auditory spatial experiences using their own graphical descriptors. However, unlike in previous studies - where elicited responses were subject to statistical and graphical analyses\(^2\) - the second part of the clarification process required that each listener describe their graphical depictions using their own verbal terminology.

The clarification study involved 20 listeners who, through their participation in previous GAL studies, were practised in representing their experiences graphically. Listeners were selected for their descriptive diversity and included individual listeners who had not previously differentiated between stimuli and others who had produced responses which did not appear to fit with the question asked by the researcher or the responses of the remaining participants. Consequently, it was essential to establish for these and the remaining participants whether: (i) a listener’s experience was at variance with their peers or; (ii) if this experience changed over time; (iii) whether using a graphical communication medium caused the ambiguities or; (iv) if a variable such as the written investigation instruction was responsible for the listener’s responses. Should either the language or - more feasibly, since listeners were using their own graphical descriptors - the investigation instruction be found to give rise to the inconsistencies in listener responses, an opportunity would exist for the researcher to misunderstand the listener’s experiences when these were being described using a graphical communication medium and the research objective would remain unfulfilled.

3.2.1 Elicitation of individual graphical descriptors

The initial part of the study - the graphical elicitation exercise - was conducted within a stationary vehicle which, although different from the vehicle used in the previous investigation, was equipped with the same multichannel audio system. This audio system up-mixed a conventional (two-channel stereo) CD for reproduction over seven loudspeakers. In this investigation, the audio system was only used in its default processed setting\(^3\).

Creation of programme material

Since the aim of the elicitation exercise was to obtain a variety of graphical descriptors from the individual listeners, 16 items of two-channel stereo programme material were created from a selection of the mono anechoic recordings available on the *Archimedes* CD. To maintain some association with

\(^2\) Graphical analysis refers to any evaluation where conclusions are based on the visual inspection of a listener’s depictions.

\(^3\) See Chapter 2, section 2.3.1 for more details of the audio system
the previous investigations, eight solo instrument and eight trio ensemble stimuli were created. Moreover, in keeping with previous investigations, the solo stimuli used the same percussive extract and sustained cello melody. Solo instruments were amplitude panned to two different locations within the eight solo instrument stereo scenes; fully left and central. Further manipulation of the solo stimuli occurred when four of the extracts were made less focused through the addition of artificial stereo reverberation. An equal amount of artificial mono reverberation was added to remaining solo instrument stimuli.

Each of the trio ensembles consisted of the same three sources as used in the initial investigations; the sustained cello melody, staccato percussion extract and female speech. Speech was in Danish to prevent participants from listening to content rather than spatial characteristics. Again, these three sources had been selected for their timbral diversity (each was distinct enough to be audible in the collective ensemble) and more critically each was selected for its spatial characteristics, ranging from the continuous nature and predominately low frequency spectrum of the cello to the transient higher frequencies of the percussive extract. These diverse sources were amplitude panned to different locations within each ensemble. Further information about the programme material is provided in table 3.2.1 below. To minimise any bias resulting from the presentation order of the programme material, six different CDs were created for the twenty listeners, each with a different order of presentation.

Table 3.2.1 Stimuli used in the clarification of listener descriptions

<table>
<thead>
<tr>
<th>Name</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Percussion Left (Perc L)</td>
<td>Percussion: amplitude panned fully left</td>
</tr>
<tr>
<td>2 Percussion Left Wide (Perc LW)</td>
<td>Percussion: amplitude panned fully left &amp; widened using artificial stereo reverb</td>
</tr>
<tr>
<td>3 Percussion Central (Perc C)</td>
<td>Percussion: amplitude panned central</td>
</tr>
<tr>
<td>4 Percussion Central Wide (Perc CW)</td>
<td>Percussion: amplitude panned central &amp; widened using artificial stereo reverb</td>
</tr>
<tr>
<td>5 Cello Left (Cello L)</td>
<td>Cello: amplitude panned fully left</td>
</tr>
<tr>
<td>6 Cello Left Wide (Cello LW)</td>
<td>Cello: amplitude panned fully left &amp; widened using artificial stereo reverb</td>
</tr>
<tr>
<td>7 Cello Central (Cello C)</td>
<td>Cello: amplitude panned central</td>
</tr>
<tr>
<td>8 Cello Central Wide (Cello CW)</td>
<td>Cello: amplitude panned central &amp; widened using artificial stereo reverb</td>
</tr>
<tr>
<td>9 Cello Percussion Voice: Narrow (CPVN)</td>
<td>Ensemble: instruments closely spaced around centre</td>
</tr>
<tr>
<td>10 Cello Percussion Voice: Wide (CPVW)</td>
<td>Ensemble: cello fully left, percussion central, voice fully right</td>
</tr>
<tr>
<td>11 Cello Percussion Voice: Left (CPVL)</td>
<td>Ensemble: cello fully left, percussion centre-left, voice central</td>
</tr>
<tr>
<td>12 Cello Percussion Voice: Right (CPVR)</td>
<td>Ensemble: cello central, percussion centre-right, voice fully right</td>
</tr>
<tr>
<td>13 Voice Cello Percussion: Narrow (VCPN)</td>
<td>Ensemble: instruments closely spaced around centre</td>
</tr>
<tr>
<td>14 Voice Cello Percussion: Wide (VCPW)</td>
<td>Ensemble: voice fully left, cello central, percussion fully right</td>
</tr>
<tr>
<td>15 Voice Cello Percussion: Left (VCPL)</td>
<td>Ensemble: voice fully left, cello centre-left, percussion central</td>
</tr>
<tr>
<td>16 Voice Cello Percussion: Right (VCPR)</td>
<td>Ensemble: voice central, cello centre-right, percussion fully right</td>
</tr>
</tbody>
</table>

Ambiguities in the previous studies led to this clarification phase
Programme material was processed by Russell Mason at the University of Surrey
Elicitation process

To participate in the elicitation phase, each listener were asked to sit in the (right-hand) driver’s seat within the motionless vehicle. When seated, individual listeners were asked to graphically describe their experience of the location of the instrument(s) present within each stimulus item, the width of solo instrument stimuli and the overall width of each trio ensemble. Before commencing the first run, each listener was asked to read a written information sheet detailing what would be required of them during the investigation. The information sheet listed the spatial attributes noted above, with a short verbal definition of each attribute. Verbal definitions used the same terminology as employed in previous GAL studies to avoid introducing any additional variability at this clarification stage. A copy of the instruction sheet is included in appendix 3.A.

Figure 3.2.1 Graphical response sheet

Listeners were provided with A3 paper response sheets for their graphical descriptions. Each response sheet featured a representation of the vehicle to aid listeners when scaling and transferring their egocentric experiences to a two-dimensional plan drawing. A blank response sheet is provided for illustration in figure 3.2.1. As with previous studies, listeners were encouraged to use their own graphical descriptive language and were informed prior to commencing the study that no correct response existed. The elicitation exercise was completed for each listener when each stimulus had been described once and eight stimuli described twice.

Rather than being analysed by the researcher at this stage, responses from individual listeners were collated in preparation for listeners to return in the second part of the clarification study and clarify their own descriptors.

3.2.2 Verbal clarification of elicited descriptions

Verbal clarification took place two weeks after elicitation to allow time for the graphical responses to be collated. After this period of organisation, all twenty listeners were required to return individually to the investigation and clarify their graphical descriptions. As with the elicitation process, clarification was accomplished with the listener sat in the stationary vehicle in order that they could audition the programme material when required. The listener was joined in the vehicle by two others (the researcher and an assistant) whose task was to facilitate discussion and independently record any terminology used by the listener. In addition to these accounts, an audio recording was made to provide a full unbiased record of each listener’s discussion. To simultaneously structure and ease the clarification process for the listener, listeners were presented with triads of their descriptions and asked to describe

---

*A full set of repeats was not possible due to the number of participants, the number of stimuli and resulting time constraints*
how two of these graphical representations were similar and yet different from a third. When conducting verbal elicitation exercises using a similar triadic presentation of stimuli, Berg - in conversation with the author - advised that obtaining verbal descriptors from listeners was not always an easy task. Indeed, Kelly himself explained that many constructs were not readily expressed by individuals using verbal terminology (1963). As the aim of the verbal clarification was for each listener to clarify their graphical descriptions in their own words, the two assistants present in the vehicle refrained from advising the listener when they had problems verbalising their experiences. Furthermore, any questioning by these assistants was limited to using only those terms already introduced by the listener. Each discussion with an individual listener continued until they were unable to articulate new verbal descriptions of their graphical responses.

---

7 This triadic method of eliciting and structuring an individual's constructs is favoured in repertory grid testing, a technique developed by Kelly (1963) and used by Berg in his work on the evaluation of spatial audio quality (2002).
3.3 Findings of the clarification study

As a result of the verbal clarification process, a tape of the discussion session and list of associated verbal terminology existed for each listener. To understand why listeners were choosing to represent their experiences as they were, it was ultimately necessary to review these verbal sources of information alongside the graphical representations elicited from each of the individual listeners. Prior to this comparison, a brief inspection of listeners' graphical descriptors was conducted to establish what was being depicted in these individual representations and where similarities and differences were occurring between listeners.

A graphical analysis of the repeated responses provided by the individual listeners indicated that most could consistently represent the same item of programme material on two different occasions. Consistency was reduced when the first presentation of a stimulus was also the first of the elicitation exercise and an inspection of listeners individual graphical responses suggested that listener 2 may have been affected in this way. When the CPV right (CPVR) ensemble was presented to listener 2 in the first run of the elicitation, this listener failed to record any information about the percussion instrument (illustrated in figure 3.3.1 above and appendix 3.B figure 3.B.9). This situation changed when the listener was presented with the same stimulus a second time (figures 3.3.2 and appendix B figure 3.B.10) and suggests the omission was a result of the investigation setting rather than the listener's experience. A second listener who appeared to develop their response style after the first run was listener 16 who used a more complex representational style in their initial description (figures 3.3.3 and figure 3.B.61 in appendix 3.B) than in their remaining depictions of the same central percussion stimulus.

The brief inspection of listener descriptions highlighted that similarities existed between listeners' representations of the percussion instruments, many listeners describing these sources as occupying a

---

8 Listener 15, one of the eleven listeners whose responses are included in appendix 3.B, was not required to describe any of the exemplary stimuli twice. Consequently consistency information is not provided for this listener

9 Exemplary graphical plots for eleven listeners are presented in appendix 3.B

10 The CPV right ensemble was created with the cello at the centre of the stereo image, the voice amplitude panned to the right and the percussion mid-way between these two instruments
narrow space within the vehicle. Exemplary descriptions of the central percussion instrument are provided for immediate reference in figures 3.3.4 - 3.3.7 with further illustrations of both percussion instruments provided in appendix 3.B. figures 3.B.11 - 3.B.13 (listener 3); 3.B.21 - 3.B.23 (listener 9); 3.B.41 and 3.B.42 (listener 13); 3.B.60 - 3.B.62 (listener 16); 3.B.80 - 3.B.82 (listener 18); and 3.B.100 - 3.B.102 (listener 20).

Figures 3.3.4 - 3.3.7 Selected descriptions of the central percussion stimulus
Figure 3.3.4 Listener 3 Figure 3.3.5 Listener 13 Figure 3.3.6 Listener 18 Figure 3.3.7 Listener 20

Although similarities in width also existed between descriptions of the narrow cello sources and the percussive instruments, a clear distinction was made between the width of the narrow central cello and the same stimulus when artificially widened. In appendix 3.B, figures 3.B.4 and 3.B.5 (listener 2); 3.B.25 and 3.B.26 (listener 9); 3.B.35 and 3.B.36 (listener 11); 3.B.54 and 3.B.55 (listener 15); 3.B.64 and 3.B.65 (listener 16); 3.B.74 and 3.B.75 (listener 17); 3.B.94 and 3.B.95 (listener 19) and figures 3.B.104 and 3.B.105 (listener 20) are all depictions which highlight the differences between both cello sources, with the widened source occupying a wider space within the vehicle that the narrower cello.

Furthermore, although the width of the percussive source (and narrow cello) was similarly represented by the majority of the listeners, less unity existed between their descriptions of the widened cello.

Figures 3.3.8 - 3.3.13 Selected descriptions of the central cello stimulus, widened by artificial stereo reverberation
Figure 3.3.8 L2 Figure 3.3.9 L9 Figure 3.3.10 L3 Figure 3.3.11 L19 Figure 3.3.12 L13 Figure 3.3.13 L15

For instance; listeners 2 (figure 3.3.8) and 11 (figure 3.B.36 in appendix 3.B) chose to represent the widened cello using two circles; listener 9 represented the same cello source as a curved box at the passenger’s seat (figure 3.3.9); listener 3 also represented the cello on the left of the vehicle but as a comparatively small source (figure 3.3.10); figures 3.3.11 above and figure 3.B.105 in appendix 3.B highlight how listeners 19 and 20 represented the cello across the front of the vehicle; whilst listeners...
Clarifying listeners’ descriptions

Chapter 3 - Development

13 (figure 3.3.12) and 18 (figure 3.B.85 in appendix 3.B) chose to describe the instrument using small central descriptors; finally listener 15 (figure 3.3.13) and listeners 16 and 17 (appendix 3.B, figures 3.B.65 and 3.B.75 respectively) chose to use a single large circle which covered the listening position and extended into the rear of the vehicle’s cabin.

In previous investigations, disparities between depictions of the same instrument could be attributed to individual listeners using different scales when representing their experiences. Here, as many of the listeners described the remaining stimuli similarly, the use of different scales was an unlikely cause of the ambiguity. It was therefore necessary to identify what else could have caused the differences in listener responses.

Since listeners were using their own graphical descriptors to describe their own experiences, it was likely that these descriptors would be meaningful to the listener and not cause any ambiguity in response. Furthermore, other than when they were first starting the investigation, listeners were not changing their descriptions over time. Thus, of the five sources of ambiguity identified at the start of this chapter, three sources of ambiguity remained plausible:

1) Differences in listeners’ auditory spatial experiences of the same reproduced audio stimulus
2) Differences in how listeners were using graphical descriptors to represent their experiences
3) A difference in how listeners were deciphering the written investigation instructions

Throughout the initial analysis of listeners’ graphical descriptions it became apparent that, although asked to concentrate only on the width and location of the different sources, with the exception of listener 3 (appendix 3.B, figures 3.B.11-3.B.20), listener 16 (appendix 3.B, figures 3.B.60-3.B.69) and listener 20 (appendix 3.B, figures 3.B.100-3.B.109), listeners favoured two-dimensional circular or box like shapes when representing their experiences of the different stimuli. Since it was expected that listeners would be concentrating on a single dimension of the reproduced events, the use of two dimensional objects in their graphical responses suggested that listeners may not have been responding as anticipated. Consequently it was necessary to turn to the listeners’ verbal descriptions to establish whether this ambiguity was a result of the listeners experiencing a reproduced event’s width differently, describing width differently, or not describing width at all. The verbal descriptors used in this clarification were extracted from the audio tapes of each listener’s discussion session. Verbal descriptor lists are provided for reference in appendix 3.C for the exemplary listeners whose graphical descriptors are illustrated in appendix 3.B.

3.3.1 Clarifying the attribute individual instrument width

When asked to verbally describe their graphical depictions of the different solo instruments 17 of the 20 listeners involved in the clarification study used the term ‘width’. That the majority of listeners happily applied this term to their graphical depictions was unsurprising as width was introduced on the written instruction sheet provided for listeners to read at the start of the study. Consequently, it was of

\[ \text{See appendix 3.A} \]
more interest to establish what was being depicted by those listeners who had not used the term width and how the term was being applied to graphical descriptions by those listeners who had. For some listeners, the term appeared changeable and could be applied to a single dimension of the reproduced sound either from front-to-back, or left-to-right in the vehicle depending on the direction the listener was facing. For others, 'width' was consistently interpreted as the 'depth' (front - back) dimension of their individual instrument descriptions and 'length' was the size of the depiction from left - right.

**Width and area**

For many listeners width was only discussed when references were made to the area covered by a particular solo source. Indeed listeners made use of multi-dimensional terms such as 'area', 'field' and 'envelope of sound' far more readily than the individual dimensions of width and depth. The prevalence of terminology dealing with an 'area of sound' and the preference listeners had shown for depicting two-dimensional objects suggested that listeners may have experienced - and therefore depicted (even when required to do otherwise) - sound as having an area, rather than an obvious width. By asking listeners to depict instrument width, it was possible that listeners were being asked to represent an attribute that was not readily experienced or easily represented using a graphical medium. Consequently, the research question may have resulted in the ambiguities in listeners' responses. Interestingly, one listener who did describe the single dimension of width by drawing a single line (listener 20, appendix 3.B, figures 3.B.100 - 3.B.109), was keen to point out after clarification that this was not what he had heard, only what he was being *asked* to depict: For instruments, when experienced, were not 'flat' but had a 'shape'.

To clarify if uni-dimensional verbal terminology (for example width, depth, height) was incompatible with a listener's graphical representation of their experiences, listeners who had used the term area were also asked to describe what they meant by this term with respect to their graphical descriptions. When describing area it was common for listeners not to mention depth; suggesting that although they had experienced solo instruments as occupying an area, this multi-dimensional area may not have comprised uni-dimensional attributes. In the pilot investigation (summarised in section 2.2. of chapter 2), listeners also appeared to give little thought to a depth attribute, however this was with respect to an ensemble and not a single instrument stimulus. When encountering a solo source, Rumsey (2002) proposed that the spatial attributes of "lateral location", "width" and "distance" were more likely to be perceived than a source’s depth and stated that the concept of individual source depth had so far proved elusive in subjective investigation. Less than half of the participants in Berg and Rumsey's initial verbal elicitation investigation (Berg, 2002, publication 1) perceived something which could be described as depth. Following subsequent analysis (Berg, 2002, publication 2) the extension of a source away from the listener and a perception of its shape were classified by the authors as "source depth". Differences in the front-back extent of listeners' descriptions were clearly illustrated by the participants in an investigation by Martens (1999) when asked to visually represent their experiences of percussive noise with varying levels of inter-aural cross correlation. However, it was Martens himself who applied the term depth to the differences in these representations. Listeners involved in the informal graphical
description stage of Martens study were instructed to record the "spatial extent", location and shape of each source rather than any specific uni-dimensional attributes of the "spatial imagery". When eliciting descriptors from participants in their descriptive analysis exercises, Koivuniemi and Zacharov (2001) found that participants agreed on the existence of "syvyyden tuntu" in Finnish, or "sense of depth" when translated into English. This was linked with whether participants could discriminate between several sound events in terms of distance. The use of depth to describe an ensemble of instruments was supported by Rumsey (2002), as depth in this respect could refer to the front-back dimension of the ensemble. Indeed, in the clarification of GAL descriptors, some listeners chose to use the term depth only when describing the distance between themselves and the instrument or instruments being evaluated or when representing the 'depth of field' (figure 3.3.14, listener 17) associated with an ensemble. A 'deep' ensemble was visualised as occupying 'more than a thin line of sound' (listener 11, figure 3.3.15) with depth for many listeners referring to situations where instruments were 'located' or 'stacked' behind one another. Similarly ensembles were described as having 'depth with respect to other instruments', where one instrument could be 'further away' or 'taking a back seat to other instruments' as demonstrated in figures 3.3.16 and 3.3.17; listeners 2 and 16 respectively.

3.3.2 Focus and localisation

When areas of instruments were described - either individually or as part of an ensemble - listeners preferred to define this area as either 'tight', 'pinpoint' or 'precise', or as 'wider' or 'broader' and filling a less 'focused', or less 'precise' 'envelope', 'pocket' or 'boundary' of sound. This widespread use of more multi-dimensional terminology indicated that area (rather than having distinct dimensions of width and depth) may have been perceived as a movement outwards from a precise core to a more 'fuzzy', less focused sound. When providing a definition of the term 'focus', Martin et al. (1999b) suggested that the use of the term differed according to the modality in use. Whereas in an optical medium focus could be considered independently of size; referring to a "measurement of the perceived definition of the object's boundaries and details" (Martin et al. 1999b, p9), focus within an audio should be related to the size of a single source. According to the authors, an image is perceived as being unfocused when a phantom image is larger or wider than the anticipated size of the source. For Martin et al., an informal discussion with their listeners identified that narrow image width, an increase in the perceptual distance to the image and less spread in the low frequency content of a source all

Figure 3.3.14 'depth of field'
Figure 3.3.15 'more than a thin line of sound'
Figure 3.3.16 'further away'
Figure 3.3.17 'taking a back seat to other instruments'
clarifying listeners' descriptions

Chapter 3 - Development

contributed to a more focused image. "Individual source width" (ISW), defined by Rumsey (2002) as the perceived lateral extent of a single source, also appears to be associated with the area covered by a single source. Rumsey explains that a source perceived as having a small ISW will be easy to locate and may be perceived as a "point source", whereas more diffuse sources - those with a larger ISW - are more likely to be poorly located.

The association between the apparent focus of a source and its ease of location was noted by listeners when clarifying their graphical representations. Terminology such as 'hard to pick out where source was coming from' was used when listeners felt themselves to be 'enveloped by sound' or the source appeared to be 'coming from all areas'. Similarly a listener could be 'bathed in sound' when they couldn't 'pinpoint' the source of the sound. Alternatively, sound could come from a 'specific sound position' a 'specific area' and have a 'fixed local image' or 'fill the whole of the area' and have a 'less focused', 'non-specific' image. In addition to being used to describe an easily localisable source, derivatives of the term 'pinpoint' were used by some listeners to describe a small source (or one opposite to a 'wide' source) with sources also being described as 'bigger' than a point. However, as large sources could be pinpointable it should be noted that pinpoint was not expressly related to the perceived physical dimensions of the source. In appendix 3.B., figures 3.B.41-3.B.50 (listener 13) and 3.B.90-3.B.99 (listener 19) illustrate two listeners' experiences of sound 'spreading' or 'projecting' away from a 'focal point' or 'main' area to cover a designated 'pocket of sound' or 'area of volume'. The concept of sound spreading away from a 'focal point' was represented by these listeners through their positioning of a letter or icon at a 'centre fill' or 'main' location of a sound and surrounding this with an 'area'. Examples include the use of small circles in figures 3.3.18 and 3.3.19 (listeners 13 and 19 respectively) and the positioning of letters by listener 17 (see figures 3.B.70-3.B.79 in appendix 3.B and figure 3.3.20 below for immediate reference).

Figure 3.3.18 'spread / projection' Figure 3.3.19 'spread across' Figure 3.3.20 'sound coming across' Figure 3.3.21 'wrapped around'

3.3.3 Width and envelopment - a continuum?

When describing their auditory experiences of the solo instruments, listeners would on occasions suggest that the sound began to 'wrap around' them, 'involve' or 'surround' them. For instance; figure 3.3.13 from earlier and 3.B.59 in appendix 3.B were described by listener 15 as 'spreading all around'; figures 3.B.75 in appendix 3.B and 3.3.14 from earlier as 'enveloping' (as if listener 17 was 'sitting amongst' what was happening); and figure 3.3.21 was described by listener 16 as 'wrapped around'.

N Ford

Doctoral Thesis
That listeners clarified their graphical descriptors using such language suggests that for some listeners, if width was being considered at all, it was being considered on the same construct scale as envelopment; with well-focused point sources opposing sounds which curved around and enveloped the listener. This is not a novel theory. One of the attribute scales constructed by a group of Finnish listeners - under the guidance of Koivuniemi and Zacharov (2001) - was called "broadness". This descriptor was defined as "how wide an area the perceived sound event seems to have" and a strong sense of broadness was equated with a sound perceived to come from all around the listener. Based on his research into spatial quality evaluation, Rumsey contends

At what point does the attribute we call source width become another one called envelopment? (Rumsey, 2002, p660)

With the benefit of graphical representation and subsequent verbal clarification, this conversion appears to occur when a reproduced source is positioned at a distance where it is possible to envelop the listener. And, since very broad sources - such as those depicted in appendix 3.B, figures 3.B.26 and 3.B.36, the voice in figures 3.B.57 and 3.B.58, the entire ensembles of figure 3.B.77 and finally figures 3.3.22 and 3.3.23 below - may remain at a distance from the listener and consequently not envelop them, it is possible to contend that envelopment is only specifically linked with broadness when there is this degree of association (in terms of distance) between listener and reproduced event.

A distinction can also be made between an event considered enveloping and one where ensemble instruments simply surrounded the listener. When describing figure 3.3.24 and figure 3.3.14 from earlier, listeners 15 and 17 used terms such as 'taking part', 'sound coming from all around me', 'sitting amongst sound', 'wrapped up amongst sound' and 'being involved' or 'enveloped'. Whereas when describing figures 3.3.22 and 3.3.25 terminology was preferred which positioned the ensemble (figure 3.3.22) or the solo sources (figure 3.3.25) at a greater distance from the listener: For instance, 'not sitting amongst the sound', 'listening to a sound' (rather than being part of it), 'not coming up to me' and 'instruments going on at a distance'.

Thus, the combined analysis of graphical description and verbal terminology suggests that not all ensembles which surround the listener are close enough to involve or envelop the listener.
3.3.4 The description of alternative attributes

Although presented with a two-dimensional response sheet showing a plan of the vehicle (without elevation information) a number of listeners represented an instrument’s position on the vehicle’s ‘z’ axis and described this position as ‘height’. During the discussion sessions it became apparent that height could be represented by placing an instrument at different horizontal positions on the two-dimensional plan. For example, a solo instrument could be depicted low by positioning it in the passenger’s foot-well and higher by positioning the source at the dashboard. Even though this appeared to be a means of solving the problem of representing height on a two-dimensional plan, a listener’s experience could be easily misinterpreted by the researcher who was without the benefit of verbal clarification; since an instrument represented on the dashboard could be interpreted either as further away from a listener or at a greater height in the vehicle than a source located in a foot-well.

Two circles were depicted by listener 2 when describing the widened central cello source of figure 3.B.5 (in appendix 3.B). When asked to describe this representation, the listener indicated that they were describing an ‘ambience’ which was distinct from the ‘main cue’ of the instrumental source. Listener 20 described their experience of the same widened cello instrument as possessing an audible ‘fill’ which occurred alongside the sound, a situation visually represented using a sweeping line alongside the shorter arrow in figure 3.3.23 (presented earlier). A third listener, listener 11 described the presence of ‘reverberation’ when there was ‘something else to the sound’ and again represented the artificially widened cello source using two distinct circles in appendix 3.B, figure 3.B.36. That listeners were depicting these additional cues suggests, once again, that individuals were not simply illustrating the width and location of the reproduced stimuli but providing descriptive information more representative of their experiences.

3.3.5 Attributes specific to ensembles

Many of the verbal terms previously identified to describe a solo instrument source were also employed by listeners when describing instruments within an ensemble. For example, the area covered by the ensemble, its precision or specificity, the listener’s ability to localise the ensemble and the possibility for the ensemble to involve or come from all around the listener. To accompany these now familiar concepts, verbal descriptors were introduced which referred to the relative positioning of the individual instruments within each ensemble. When instruments within ensembles were depicted over the width of the vehicle (for example: figures 3.B.18, 3.B.47 and 3.B.97 in appendix 3.B and figures 3.3.26 and 3.3.27 overleaf) these graphical responses were verbally described as having ‘well defined instruments within space’, or containing individual instruments which had their own ‘well localised’, ‘separated’ ‘image’ or ‘envelope’ within the ensemble. Conversely, figure 3.3.28, and figures 3.B.37, 3.B.76 in appendix 3.B, represent situations where listeners could not differentiate between the various instruments within the ensemble. Here instruments were described by listeners as ‘joined’, ‘merged’, ‘blended’ or ‘mixed’ into one sound with the same boundary. Figures 3.B.38 to 3.B.40, 3.B.56 and 3.B.57 illustrate conditions where instrument boundaries were perceived to ‘overlap’ within the
Clarifying listeners' descriptions

ensemble. In these situations, even though sources were positioned very close together, listeners felt able to define these instruments as 'distinct sounds'.

For many listeners, the point at which a sound occupied a distinct space within the vehicle was the instance when it was said to occupy a different 'ensemble'. Consequently for these listeners, the original ensemble of three instruments could itself be split into further ensembles according to whether or not instruments occupied the same space. For example, figure 3.3.29 could be interpreted as containing two 'ensembles' whilst figure 3.3.28 and figures 3.B.37 and 3.B.76 in appendix 3.B could be described as one single 'ensemble'. The division of a reproduced scene into several ensembles was discussed by Rumsey (2002), who suggested that several levels of "ensemble width" may be required to describe the characteristics of "groups within groups". However, ambiguities may still occur when trying to establish group membership for the different instruments. For instance, although located in the same position, does figure 3.3.30 (and figures 3.B.66, 3.B.96 and 3.B.106 in appendix 3.B) indicate one or more distinct groups? Are figures 3.3.31 and 3.3.32 (and figures 3.B.50, 3.B.69 and 3.B.109 in appendix 3.B) indicating two or three ensembles? And does the large envelope around the circles in figure 3.3.24 (from earlier) and similarly the large circle around the letters in figure 3.3.33 suggest these should be read as a single ensemble rather than as three distinct sources? It is clearly possible to misinterpret a listener's auditory spatial experiences at this point.
3.4 Summary of the findings from the clarification process

In previous investigations, the communication of a listener’s auditory spatial experiences through individual graphical representation resulted in ambiguous responses. The aim of the clarification process was to establish why these ambiguities were occurring in order that a step could be taken towards fulfilling the research objective, namely: To develop a descriptive graphical language which enables the structuring and representation of listeners’ auditory spatial experiences in order that these experiences may be understood when communicated to the researcher.

When individual listeners clarified their graphical descriptions using verbal terminology, it was identified that the verbal attributes of width and location were either being applied to their experiences in a manner other than expected, or listeners were avoiding describing the width attribute altogether, appearing to answer an alternative question to that asked by the researcher. Koster notes that:

Whenever one puts a question to a subject, one always gets an answer, even if the question is in principle unanswerable (Koster, 2003, p7)

Support for the existence of an unanswerable question is evident in listeners’ preference for two-dimensional graphical shapes and the employment of two-dimensional verbal terminology when asked to describe their experiences. Further support is apparent in the use of verbal descriptors unrelated to width (for example ambience and height), and the specific statement by one listener that he was responding to the research question rather than representing his actual experiences.

Thus it may be that the uni-dimensional descriptive term width cannot be applied appropriately to a listener’s experience when this is to be represented graphically. Listeners themselves may merely perceive, and therefore depict sound as occupying a two-dimensional area or three-dimensional volume of space within a vehicle. The area may be well defined or less focused and likewise, instruments within this area may be easily located at a specific position or more diffuse. It is also possible that the listener perceives an ‘envelope of sound’ for each distinct instrument or group of instruments, the sound ‘projecting’ or ‘spreading’ away from the ‘focal point’ or ‘centre’ of this envelope. If, as it appears, a listener graphically represents sound in terms of more multi-dimensional attributes, asking a listener to graphically describe a uni-dimensional verbal attribute could result in: (i) the ambiguous attachment of this attribute to their experience; (ii) a listener’s inability to identify any experience which relates to the verbal attribute or; (iii) the listener ignoring the request of the researcher and focusing instead on describing attributes representative of their actual experiences. Aside from identifying the consequences of asking an unanswerable question, it is also worth considering why listeners in previous GAL studies were being asked to describe the width of an auditory spatial scene.

The decision to ask listeners to describe the width of an ensemble was made following the findings of the pilot investigation. As mentioned in chapter 2, the pilot study provided listeners with an exemplary descriptive style to use, should they wish, when graphically describing their experiences. Other than
this provision, the listeners were not limited by a specific research question detailing which experiences to represent - the only obvious restriction being the A4 tracing paper used as a response sheet. When graphical responses were measured and quantitatively analysed, it was found that differences could be identified between the different left-to-right extents of the various complex musical stimuli. Considering width (or a differently named attribute with similar characteristics) has been - and continues to be - commonly elicited as a verbal descriptor of both ensemble stimuli and individual instrument sources\(^{12}\), no further confirmation was sought regarding the suitability of either descriptor for graphically eliciting from listeners their experiences of a reproduced audio event. Only now, following clarification, is it possible to contend that width, although an effective verbal descriptor of a listener’s experience and a useful quantitative measure of a listener’s graphical representations, ought not to be specifically linked with a listener’s experience when this is to be represented graphically.

In conclusion, since an individual’s own language is meaningful to them, a listener’s communication of their experiences within their own descriptive graphical language can be accepted as valid if what is being communicated is a representation of their experience and not an external factor (such as one introduced by the investigation setting). The clarification process has highlighted that posing an inappropriate question - one that fails to match with a listener’s experiences - is such a factor and can result in a listener either ignoring the question, or (more problematically) ignoring their experiences. If the researcher does base their comprehension - and any further descriptive communication - on a description which fails to represent the listeners’ experiences, there are obvious problems for the validity of this understanding and any ensuing descriptive account. Likewise if a listener chooses to respond to an alternative question, one which better matches their experiences, although the listener’s representations will communicate their experiences, the researcher’s comprehension will reflect the question and (once again) the objective of the research remains unfulfilled. A graphical representation of this explanation is provided in figure 3.4.1: GAL development model stage II\(^{13}\).

\(^{12}\) For a full description of the verbal attributes ensemble width and individual source width, plus details of their elicitation and use in subjective audio research, see Rumsey (2002)

\(^{13}\) A key for the descriptive process model is provided in appendix 1.
In figure 3.4.1 ‘Q’ is used to refer to the written instruction provided at the start of the investigation by the researcher; ‘St’ denotes the specific reproduced stimulus; ‘Exp¹ - Expⁿ’ the possible auditory spatial experiences of a listener; ‘IR’ a listener’s individual representations and; ‘C’ refers to the researcher’s comprehension of the listener’s experiences via this representation. Ambiguities in communication are indicated by a dotted line.

Figure 3.4.1 GAL development model stage II: Known ambiguities following the clarification of listener descriptions

- The researcher misunderstands (CQ) the listener’s experiences because the listener represents (IRQ) the researcher’s written instruction (Q) and not their experiences (Objective not fulfilled).
- Listener uses own graphical descriptors to represent their experiences in (IR¹). Researcher has potential to understand listener’s experiences (CR¹) (Potential for research objective to be fulfilled).
- The researcher misunderstands (CM¹) the listener’s experiences because the researcher does not fully understand what the listener is describing using their individual representations (IR¹). (Objective not fulfilled).
- The researcher misunderstands (CM²) the listener’s experiences because the listener represents their experience (IR²) rather than the researcher’s written instruction (Q) (Objective not fulfilled).
3.5 Rationale for the continued development of GAL

Using figure 3.4.1 as illustration, it is clear from this model that ambiguities in graphical responses can be minimised by the careful selection of the terminology used in a written investigation instruction, ensuring that this terminology better reflects the actual experiences of the individual listener. One outcome of the clarification process was a set of verbal descriptors, a word list, for each listener (examples of which are provided in appendix 3.C) which could ostensibly be used by the researcher to communicate with the listener in a language more reflective of their auditory spatial experiences and of the graphical descriptors created to represent these experiences. Although two assistants made note of the verbal descriptors used by the individual listeners during the clarification process, only one assistant (the researcher) extracted terminology from the audio tapes of each listener’s discussion sessions for the descriptive account and individual word lists. Since no additional validation of either the account or the word lists was conducted, it was not possible at this stage to say whether the verbal terminology extracted by the researcher was truly appropriate for the communication of a listener’s experiences. Furthermore, as noted by Kelly in his *Psychology of Personal Constructs* (1963), even if the question asked by the researcher in future investigations was able to reflect the listeners’ experiences by using verbal terminology collated from the clarification process, there would still exist the possibility that each listener would choose to use their individual graphical descriptors differently when responding to this appropriate question. As Kelly asserted:

> Even those constructions which are symbolized by words are not necessarily similar just because the words are similar. Conversely, two persons may be using essentially the same constructions of their experience, although they express themselves in quite different terms. (Kelly, 1963, p92)

Consequently (as illustrated by the outcome CM' in figure 3.4.1), even after the clarification process, there still existed the potential for the researcher to misinterpret a listener’s depiction of their experiences when this representation used individual descriptive language. To establish whether the verbal terminology extracted from the taped discussion sessions by the researcher actually reflected the auditory spatial experiences of the listeners, it was decided to return verbal descriptor lists and graphical representations to listeners for further description and validation.

3.5.1 The communalisation of experiences

Although it is crucial for any description of a phenomenon to be initially undertaken by the individual - so that this description reflects the phenomenon (the stimulus) as experienced by the individual - a more thorough understanding of the stimulus as experienced can be obtained if this individual description is shared with others. “Communalisation” or “inter-subjectivity” is possible because, as Köster asserts, even though each individual will have their own personal history, “people may have much more in common in the way they experience situations than it seems at first sight” (Köster, 2003, p370). Köster consolidates his assertion by stating “if situations were really completely individual and private, communication of feelings would be impossible. Literary books would not be published and
the film industry would not exist" (ibid.). Köster exemplified inter-subjectivity as the ability of an individual to inquire of someone else the identity of an unknown object; a task which is only possible since the individual is "absolutely sure that he [the other person] sees the same shape of the object". Köster continues by differentiating between recognising the same physical object and applying a universal meaning to the object; meaning being applied independently even when an object is universally recognised. For example "the fact that [the other person] notices something that you have overlooked although you can clearly see it when it is pointed out to you, is because he also sees meaning" (ibid. p368).

Although Kelly's *Psychology of Personal Constructs* is based on the 'individuality corollary' that an individual construes events as experienced by themselves, the author posits what at first appears to be a contradictory 'commonality corollary'. Commonality is defined: "to the extent that one person employs a construction of experience which is similar to that employed by another, his psychological processes are similar to those of the other person" (Kelly, 1963, p90). Thus, for Kelly, it is the construction of an experience that provides the basis for similar action and not the identity of the experience. Accordingly, both commonality and individuality corollaries are free to co-exist. Kelly clarifies his position stating:

> Experience, as we have defined it, is a matter of successively construing events. To construe experience, then, is to take stock of the outcome of this successive construing process. Thus, if two people take similar stock of their successive interpretations, their behavior will exhibit similar characteristics. The historical development of their thinking need not be similar – only the stock-taking need be similar. Hence it is not the similarity of experience which provides the basis for similarity of action, but similarity of their present construction of that experience (ibid. p92)

Kelly is adamant in his belief that individuality in experience does not rule out the possibility that events can be construed similarly: "To say that two things differ from each other in every conceivable respect is to express the ultimate in particularism and to leave one's listener in a confused state of mind about the whole matter." (ibid.)

For Civille and Lawless, commonality is required in the way individuals describe external stimuli. Here, the discussion of terminology and underlying perceptual structures has benefits for panelists involved in the development of a descriptive language which include; "the ability of individuals to monitor and check the inferences and reasoning of the others in the group which then speeds the process of concept attainment by all members" (Civille and Lawless, 1986, p211). According to Moustakas, "in the back and forth of social interaction the challenge is to discover what is really true of the phenomena of interpersonal knowledge and experience" (Moustakas, 1994, p57). Consequently it is believed that an individual's knowledge may be influenced or "corrected" by looking at the phenomenon "from the perspective of another self" (ibid. p94); this social interaction leading to the adoption of a more inter-subjective description of the phenomenon by the individual. For Moustakas:
Following our own self-evidence of what appears to us, we check with others regarding what they perceive, feel, and think. In the process of this kind of careful checking we may revisit the phenomenon and discover something new that alters our knowledge of the thing (ibid. p95).

Husserl, cited in Moustakas (ibid.) stated that this process of communalisation alters validity through "reciprocal correction"; enabling each individual to obtain a more complete understanding of the phenomenon being revisited. Similarly, for Kelly it is the constant exposure to an event that causes the individual to challenge the constructions which are already in existence.

Successive revelation of events invites the person to place new constructions upon them whenever something unexpected happens. Otherwise one's anticipations would become less and less realistic. The succession of events in the course of time continually subjects a person's construction system to a validation process. The constructions one places upon events are working hypotheses, which are about to be put to the test of experience. (Kelly, 1963, p72)

Consequently it was decided that individual graphical representations and descriptive verbal terminology would be returned to small groups of listeners for further inter-subjective description and discussion. The objective of this group discussion was twofold; firstly; to facilitate a more comprehensive understanding of listeners' auditory spatial experiences by both listener and researcher (improving the researcher's ability to provide a valid descriptive account of these experiences); and secondly, to develop - through the consolidation of the individual word lists and graphical descriptors of the individual listeners - an inter-subjective (universal) graphical language appropriate for the representations of each listeners' experiences and a comprehensive set of verbal terminology for use in any verbal communication between researcher and listener.

The decision to create a universal graphical communication medium - based on listeners' inter-subjective knowledge of a phenomenon - was made to rectify the problem highlighted by CM1 in figure 3.4.1. Namely that when a listener uses their own individual graphical descriptions to represent their experiences the researcher may not fully understand what is meant by the descriptors and in this way may provide an unrepresentative descriptive account of a listener's experiences. By creating a graphical communication medium which could be used by all listeners to represent their experiences it was envisaged that the researcher would be able to base their explication on the listeners' agreed application of graphical descriptors to experience, rather than having to identify how each listener was using their own unique descriptors to represent their own experiences.

Thus, through the process of returning an individual listener's graphical and verbal descriptors to small groups of listeners, it was envisaged that known sources of ambiguity in communication from researcher-to-listener (the verbal terminology of the investigation instruction) and listener-to-researcher (the listener's use of individual graphical descriptors) could be minimised.
3.6 Developing a universal set of graphical descriptors: Part 1

The procedure for developing the descriptive universal graphical language (U-GAL), and associated verbal descriptors, conformed in part to the language development process of Quantitative Descriptive Analysis (QDA) outlined by Stone in Hootman's *Descriptive Analysis Testing* (1992). The applicability of this process to subjective audio evaluation has previously been highlighted in Chapter 1 of this thesis, with an overview of the method presented by Bech (1999) and subsequently employed by Zacharov and Koivuniemi (2001a) in their development of the *Audio Descriptive Analysis & Mapping* procedure (ADAM). For Stone, the development process was one of panel training in which "panelists, as a group, meet with the panel leader and develop a common language that describes their perceptions of the products" (Stone, in Hootman, 1992, p17). Although seen as a training process in QDA - with panelists creating a language to be used when describing the perceived attributes of external reference stimuli - the same process of group meeting and discussion could be used to create an inter-subjective descriptive language capable of representing listeners' experiences on auditioning an external stimulus.

3.6.1 Language development process

The validation of verbal terminology and development of the U-GAL required the participation of the same 20 listeners who had provided graphical descriptors and verbal accounts of these individual descriptions in the clarification study. In order that even the most reticent communicator would have a forum for discussion, listeners were divided into five small groups - a practice supported by Koivuniemi and Zacharov (2001). To promote discussion, each panel consisted of listeners with varying representation styles. To minimise any bias originating from placing a group of enthusiastic experienced listeners with an introverted novice, the previous listening experience of each individual - and their willingness to communicate in earlier studies - were also taken into account when listener's were allocated to a panel. Individual discussion sessions were timetabled into 90-minute slots. However this time limit could be extended should panelists become involved in intense debate.

The task of each panel

When developing a descriptive language for use in the representation and subsequent evaluation of external stimuli, Meilgaard et al (1999) outlined the role of the panelists as follows:

1: Panel generates original list of descriptors to which all panelists are invited to contribute
2: Original list, containing many overlapping terms, is rearranged and reduced into a working list in which the descriptors are comprehensive (describe the product category completely) and yet discrete (overlapping is minimised)

The task of each four-person panel involved in the development of the universal graphical and verbal languages was to construct two sets of descriptors (one verbal, one graphical) which could be used by the whole panel to represent their experiences. Individual word lists and graphical representations already existed for each panel member (as a result of the clarification process). Consequently, the descriptors for all listeners on a panel were made available to that panel from the outset, and step 1
from the above list was replaced by a process of each listener revisiting and discussing with the remainder of their panel the auditory spatial experiences they were trying to represent with their graphical and verbal descriptors. Since verbal descriptors had been selected from taped discussion sessions by the researcher and not the listener, panelists were also instructed to remove words from their list which they considered to be of little relevance and to add words to their list where their experiences had not been fully captured by the researcher's description. Towards the end of this initial discussion phase, descriptors for each panelist were grouped according to similarities in the experience being represented.

The initial explanation of experiences and descriptors by individual panelists led automatically to the discussion between listeners of possible overlaps in their application of terminology to their experiences. Panelists used this opportunity to further clarify their descriptive terminology, establishing whether the same terminology was being used by different members of the panel to represent similar experiences or if differences existed in their application of descriptor to experience. Following a lengthy discussion period - which often extended past the allocated 90 minute session - panelists attempted to consolidate their individual terminology into universal graphical and verbal languages. To be included in a panel's language, a descriptor had to meet the criteria that it could be used by all panel members to comprehensively describe a distinct spatial attribute of a (selected) reproduced audio event as experienced by them. In addition to developing both graphical and verbal descriptors, panels were asked to indicate how differences in intensity could be described using their descriptors, for example: how would a focused area be communicated graphically and how would this compare with a less focused area?

Role of the panel leader
Each discussion session took place under the supervision of the researcher, whose role as 'panel leader' is defined by Stone as follows:

The panel leader facilitates the discussion, ensures that materials needed by panelists are available, keeps notes, but does not participate in the actual development of the attributes needed to fully describe the products. (Stone, in Hootman, 1992, p17)

To assist in the description of their experiences, panelists were provided with blank A3 graphical response forms, the same as had been used in the clarification process and illustrated previously in figure 3.2.1. To further assist in their development of the graphical and verbal languages, the vehicle which had originally been used in the clarification study was also made available to listeners in order that they may re-audition the programme material which had prompted their initial experiences and subsequent descriptions. For the practical reason of limited space, discussion sessions themselves did not take place within the vehicle. Discussions were audio taped and transcribed (where possible between meetings) to provide panelists with an on-going record of their progress and to maintain an unbiased account of panel meetings.
In accordance with Stone’s defined role for a panel leader, the researcher did not participate in the actual development of the graphical or verbal descriptors, but was available to provoke discussion and clarify the terminology being used by the panel. The involvement of the researcher was limited to using only the language of the panel to avoid biasing the development process. This is contrary to Stone’s later description of the role, which states that a panel leader *may* become involved in the development process and suggest attributes “when the panel is experiencing difficulty describing a particular sensation”, with the proviso that “the panel, as a group, must come to agreement as to inclusion of each attribute” (ibid.). As a further precaution against the panel leader biasing the language development, an additional (non-participating) panelist was present throughout the group sessions to monitor the contribution of the panel leader and to provide technical support.

The time taken by each panel to develop their graphical and verbal descriptors varied: Two panels each required a total development time of around three hours to consolidate their individual descriptors into a language which was mutually agreed upon by all panelists. A further two groups consolidated their graphical and verbal terms in four approximately 90-minute meetings, with the final panel requiring five roughly 90-minute sessions before finally arriving at a verbal and graphical language which was acceptable to all.

### 3.6.2 Identifying commonality in experiences: Précis of panel discussions

During each panel’s first meeting, panelists were re-acquainted with their graphical descriptions and introduced to the individual lists of verbal terminology constructed by the researcher as a result of the clarification process. The period of explanation and refinement which ensued resulted in panelists not only clarifying and structuring their individual descriptors into more coherent lists, but identifying (as a panel) the common experiences requiring representation in their graphical and verbal languages. Since each panelist’s individual list of verbal terminology was divided into two separate sections - with descriptors representing solo sources separated from those used by individuals to represent trio ensembles - panelists were required to structure their experiences in such a way as to develop descriptive terminology which could be used to represent both distinct source types. Consequently panelists were required to identify the exact constituents of an ‘audio image’.

#### Deconstructing the audio image

When discussing how to describe their experiences of an audio stimulus (which could be either an ensemble or a solo instrument) similarities existed between panels in how their auditory spatial experiences were being structured and described. One panel decided that any stimulus could be best represented as a ‘soundstage’ or ‘audio image’. An ‘audio image’ was defined as an area which could be filled by any number of individual instruments occupying the same location within the vehicle. For the panel there existed the possibility that more than one audio image could exist simultaneously within a vehicle; should the same instrument be perceived to divide and occupy two different locations in the

---

14 Exemplary graphical descriptions and verbal terminology for individual listeners are provided in appendices 3.B and 3.C respectively.
vehicle, or a gap exist between instruments within the same ensemble. A second panel used the term ‘soundstage’ and one again cited the location of the different instruments within the ensemble as their means of differentiating between the different soundstages which could exist for the same audio event. One further panel used location as a means of deconstructing the audio stimulus, maintaining that when several ‘sounds’ were perceived as coming from one place a listener could just sketch these as having a combined, integrated ‘field’: However different fields should exist for auditory events at different locations. Relevant extracts from the transcripts of the taped discussion sessions for these three exemplary panels are provided in appendix 3.D: excerpts 1 - 3\textsuperscript{15}. By deconstructing an auditory event into distinct parts according to stimulus location, panelists from the above panels believed they would be able to describe each distinct soundstage, audio image or field in terms of its characteristics. These characteristics are outlined below.

**Audio image position**

After much debate, panels eventually decided that only the front-to-back and left-to-right position or location of an audio image, field or soundstage could be described graphically; due to the physical restrictions of the two-dimensional response sheet. However many panelists did discuss their experience of sound being located in three-dimensional space, with this third dimension being described as the ‘height’ of an event. Extracts from transcripts of two panels are provided in appendix 3.D (excerpts 4 and 5) and highlight how height was being described during the language development sessions. It should also be noted that many panelists suggested the creation of a three-dimensional response form to better enable the representation of this important elevation dimension. Further to describing the location of an audio image, field or soundstage, panels discussed how a sound could be positioned in relation to a listener. One panel believed it was possible to feel ‘amongst’, ‘enveloped’ and ‘involved’ by a sound when ‘surrounded’ by what was going on; a panelist from this panel stating that this was ‘the difference of sitting listening to someone playing a piece of music and the difference of actually being amongst the people playing’. The same panel distinguished between the experiences of being ‘surrounded by’ and ‘close to’ a source: It being possible for a listener to be close to an event without feeling part of (or ‘involved’ in) it. A second panel described a feeling of ‘envelopment’ as the feeling that ‘you are in the centre of the sound - sound is coming from all around’, with the possibility that a listener was either ‘involved’ with the image, or ‘not involved - outside of the image’.

**Size and area**

Many panels debated how best to describe experiences typically represented by the terms ‘area’ and ‘size’ using graphical descriptors and verbal terminology. It was concluded by four panels that area consisted of two-dimensions (width and depth) which could also be described independently. Much debate was required by one of these four panels - as exemplified in the transcript extract (excerpt 6) provided in appendix 3.D - before agreement could be reached as to whether the depth descriptor could be applied to panelists’ individual experiences. Within the panel, depth represented either the front-to-back.
back axis through an elliptical area of sound, or a measure of 'projection'; the distance from a 'centre of sound' to its forward boundary. Another member of the same panel perceived 'sound' as having no associated 'depth' and consequently their only use for the depth descriptor was in representing the distance between the source of the sound (the loudspeaker) and the perceived location of the event. After much discussion, the panel decided that area could be described as a two-dimensional attribute ranging from 'small' to 'large', the attribute requiring no further representation (other than to describe the location where this area was experienced). For the fifth panel, although area was discussed in terms of its width, no mention was made of depth during the panel's discussion sessions. Instead the panel decided that their experiences could best be represented by a distinct width descriptor and the 'area the sound filled'; from 'small (pinpoint)' to 'large (whole car)'.

Source localisability
When discussing the ease with which they could locate their experience of the different audio events, four panels described the presence of an 'origin', 'focal point', 'main part' or 'centre of sound': The fifth panel describing an apparently similar experience using the verbal descriptor 'localisability' to distinguish between sources which were easy to localise, difficult to localise and those which couldn't be localised. For all panels, descriptors were used to represent the focal point of any localisable source within a reproduced stimulus and could therefore exist for both solo instruments and any of the individual instruments within an audio image, field or soundstage. Around the 'origin' or centre of a localisable source, panelists from three of the panels described an 'area' which was 'filled' or 'covered' by the same sound. In one of these panels, the 'origin' could change size within the 'area' according to whether it was perceived to cover the whole vehicle or a specific point. For the other two panels, a 'focal point' or 'centre of sound' only existed when the source was 'definable' or 'focused': Should a 'centre of sound' be definable or 'pinpointable', one panel then specified the position of this definable centre. Although the fourth panel did discuss the existence of a main part, not all members of this panel surrounded this main part with an 'image' or 'shape' which denoted the outer perimeter of the source. Instead the surrounding area was described by one panelists as an 'area of confusion' and was linked with how easily the panelist could locate the sound: The area increasing in 'blurriness' or 'fuzziness' as an instrument's location became less precise and more difficult to ascertain. Although the panel eventually excluded 'area of confusion' from their descriptive language as a result of its ambiguity (it being impossible to differentiate between a large graphical description representing a 'confused' image and a similarly sized description which represented a physically larger area) the panel was not alone in likening a small area to a 'focused', 'certain' or 'concentrated' sound and a larger area with an 'unfocused', 'dispersed' event.

Excluded descriptors
Throughout the process of developing their descriptive languages, panels were encouraged to exclude individual terminology which could not be used by all panel members to describe their individual experiences. In some cases discarded terminology was replaced by a descriptor which could be used to
represent panelists' experiences. However, if an actual auditory spatial experience was deemed too obscure, the experience itself would not be represented in the panels' languages.

Ambiguity and lack of consensus as to how 'area of confusion' could be applied to a listener's experiences led to the exclusion of this descriptor from one panel's language. Within the same panel, 'hugging the space' was again excluded from the verbal language as not all panelists were able to associate this descriptor (for a sound which 'clung to the boundaries of the vehicle' as a result of it lacking in 'depth') with their experiences. In a separate panel, 'wall of sound' was used by one panelist to represent a large, 'unfocused' elliptical sound which could exist either in front or behind the listener and 'bathed in sound' was similarly applied to an experience which included the listener. Both terms were excluded from the panel's verbal language as panelists believed the descriptors represented the same experience as an 'area of sound' without a definable 'centre of sound'.

The verbal descriptors 'image shift' and 'sound stability' were introduced by two different panels. Both terms were used to describe a moving source within the vehicle, but panelists could not develop a means of graphically representing this experience in a comprehensive manner and the two descriptors were therefore omitted from their languages. The descriptor 'point source' was omitted from one panel's verbal language as panelists believed their experiences could be better associated with the verbal descriptor 'easy to localise'. For the same panel, 'soundstage' - the term initially applied to the distinct groups of instruments created from a reproduced audio event when this was deconstructed according to instrument location - was eventually replaced by 'audio image'. Soundstage was considered too technical a term by some panelists; others could use the term to describe the qualities of a physical stage but felt it had little bearing on their perception of a reproduced stimulus; and, for the remaining panel members, the term did not represent anything. Panelists believed 'audio image' was more reflective of their actual experiences.

Many panelists suggested the inclusion of a descriptor in their languages which could be used to represent the relative loudness of the individual instruments within an ensemble stimulus. One panel - close to including a representation of loudness in their graphical language - originally suggested this be denoted by changing the thickness of graphical width and depth descriptor lines, but finally decided to develop only verbal terminology. A second panel noted that it would be useful to rank individual instruments within an audio image according to their loudness. However, the remaining panels omitted loudness from their languages as they did not consider it to be a spatial attribute.

Ambience and ambiguity

Two panels concerned themselves with discussing the concept of ambience. The term was used by panelists to represent any experience which could also be described as a 'feeling' accompanying any sound: A 'feeling of space' or a 'feeling of experiencing a live sound'. Although the experiences behind the terminology were relevant for all panelists, the actual descriptor 'ambience' was considered by both panels to be far too ambiguous to be including in their verbal languages. Extracts from the
transcripts of the panels' discussion sessions are included in appendix 3.D (excerpts 7 - 10). These extracts highlight the complexity of the term ambience and the disparate experiences being represented by the descriptor. Ambience was ultimately dropped from both panels' languages in favour of the term 'feeling of space' which represented the 'spaciousness', 'size of the space' or 'the environment a listener felt themselves to be listening in'. A second distinct descriptor was also created by one panel for the 'feeling of experiencing a live sound'. Although two panels discussed their experiences with reference to a feeling of space only one panel developed a graphical descriptor for this experience - an irregular cloudy shape - as the other panel was unsure of how to represent such a feeling (see excerpts 7 and 8). Other experiences were more successfully converted into graphical descriptors by panelists.

### 3.6.3 Converting experience to a graphical language

As previously identified, panelists deconstructed reproduced audio events into 'audio images' consisting of solo instruments or groups of instruments perceived to occupy the same location or area within the vehicle. Thus any number of audio images could exist within the vehicle and panelists were tasked with developing a graphical language which could reflect this possibility. Although the location of an audio image did not require the development of a graphical descriptor - as panelists were able to indicate the position of an image (and their involvement with this image) by placing descriptors on a response sheet - other experiences required the development of specific graphical descriptors.

As the concept of sound occupying an area was common to all panels, a graphical representation of this experience was amongst the first to be developed. Most panels decided that area could be depicted using different sized boxes, circles, or 'race-tracks' (boxes with rounded corners). For another panel, the decision was made to describe an instrument's shape using front-back and left-right arrows. By dividing each instrument into these component dimensions the panel felt they would be able to describe any shape experienced; as instruments did not necessarily possess a conventional outline which could be described using a rectangle or circle. One further panel did use a circular outline to denote the boundaries of an image but also used arrows to determine the shape of the sound being delineated. Although the panel that used arrows to represent the shape and size of a reproduced stimulus employed different colours with these arrows to represent individual instruments, most panels decided to represent separate audio images using separate area descriptors. Two panels suggested that individual instruments at the same location should be represented by placing numerous area descriptors directly on top of each other. However, other panels employed a single area descriptor in these circumstances. As multiple instruments could exist within the same area, panelists recognised the need for identifying each instrument. Most panels believed this could best be achieved using a denotative letter for each instrument occupying an area.

The 'focal point' or 'origin' of each localisable source within an audio image also required representation by a graphical descriptor. In developing an appropriate descriptor for this experience, four different representations were proposed by the panels: a circle; a letter (which was also used to identify the type of instrument); a combination of letter and circle; and an 'x'. The group who believed
origin could change size created a scale of four different sized circles to represent this attribute. Other panels stated that either the descriptor was present when a sound was localisable, or removed from a listener’s depiction when there was no focal point. Once again - as multiple instruments could exist within the same audio image - there existed the possibility that multiple focal points could exist within the same area when this consisted of more than one definable instrument.

Thus to describe a single audio image in the simplest terms of its area and focal point five similar graphical lexicons existed; one for each panel. These graphical languages are illustrated in Figure 3.6.1.

Figure 3.6.1 Graphical representations of area & focal point as developed by the five individual panels
3.7 Developing U-GAL part II: Consolidating panel languages

At this stage of the language development process, five different graphical languages and associated sets of verbal descriptors existed - one for each panel. Further consolidation of these five languages was required to arrive at a universal graphical language (U-GAL) which could be used individually by all panelists to describe their auditory spatial experiences. The development of a universal set of verbal descriptors for describing panelists' experiences was also required in order that the researcher could (i) use appropriate language when providing verbal information to panelists and (ii) provide a descriptive account of a listener's experiences.

Five panelists - one per panel - were selected by their peers to discuss their panel's languages in a final consolidation process. The process for developing the universal graphical language and associated verbal descriptors followed the same descriptive analysis format used when the panel languages were developed from listeners' individual word lists and graphical depictions. However, panelists were this time presented with a précis of each panel's individual discussion sessions and list of their graphical and verbal descriptors. The five remaining panelists were involved in three discussion sessions of approximately 90-minutes each, where they debated which of their descriptors most comprehensively described a listener's experiences and would therefore be included in the U-GAL.

3.7.1 The universal graphical language (U-GAL)

After much discussion about the merits of each panel's individual graphical languages, panelists decided that a listener's experience of a single instrument could be thoroughly represented using four complementary graphical descriptors as illustrated in figure 3.7.1.

Figure 3.7.1 Graphical language for describing a single instrument

Representing a Feeling of Space

Within U-GAL, the cloudy shape illustrated in figure 3.7.1 was developed to represent a 'feeling of space'. Although panelists provided no explicit verbal definition of the experience being represented by this descriptor, it was linked in discussion with the 'size of the environment' the listener felt themselves to be listening in. Panelists suggested that the environment could range in size from a small room (for example a pub) to a large space such as a cathedral or arena. It was also possible for no 'feeling of space' to exist and in this situation the descriptor could be omitted from a listener's graphical representation. Since it was possible for different environments to be experienced by a listener, the
‘feeling of space’ descriptor was developed to change size from the size of a seat within the vehicle to the size of the vehicle itself. However, at this stage it should be noted that even though all panelists agreed to the inclusion of the ‘feeling of space’ descriptor in the graphical language, one panelist had reservations about how listeners were actually going to use the descriptor: Since listeners were actually sat in a physical environment - the vehicle - with its own dimensions, the panelist was unsure how listeners were going to differentiate between the actual space of the vehicle and the environment they felt themselves to be listening in. A second panelist voiced concerns about the researcher’s ability to accurately describe a listener’s representation which was without a ‘feeling of space’ descriptor: How could the researcher differentiate between the listener not experiencing a ‘feeling of space’ and simply forgetting to use the descriptor? These concerns were left unresolved at the end of the panel discussion sessions; panelists believing they would be settled through the process of validating the language.

Representing the Size of Sound
The ‘size of sound’ descriptor (the round-cornered rectangle in figure 3.7.1) caused few problems for the final panel since this concept had been described by all panelists in previous discussion sessions. The ‘size of sound’ descriptor was developed to represent the size of the area filled or covered by each separate audio image. After much discussion - and wary of how instruments could change in shape as well as size - the panel decided that the listener should be able to expand the ‘size of sound’ descriptor in any direction in order to represent their experiences comprehensively.

Representing the Centre of Sound
The ‘centre of sound’ - an interesting misnomer since this descriptor could be positioned anywhere within a ‘size of sound’ - was represented graphically using a small circle. Although panelists agreed on the actual graphical descriptor and the verbal term ‘centre of sound’ (with its accompanying definition of ‘core of sound; the main focal point of the sound’) the application of this descriptor to a listener’s experience was the subject of considerable debate for the panel. The discussion mainly centred around whether a listener would wish to represent a difference in the perceived magnitude of the ‘centre of sound’. One panelist stated unequivocally that his panel had required a distinction to be made between an ‘origin’ which occupied a ‘point source’ (originating for example, from the ‘left front of the vehicle’); an ‘area source’ (such as the whole front of the vehicle); or the ‘whole vehicle’. Yet the remaining panelists were content to describe a ‘centre of sound’ as either present or absent: Present when a source had a ‘localisable’, ‘focused’ or ‘definable’ ‘core’ or ‘centre’ and absent when the source was less easy to localise. A further contention with ‘centre of sound’ was how to graphically indicate when an auditory event was not localisable? Although most panelists agreed that this could be described by a listener omitting the graphical ‘centre of sound’ descriptor from their representation, it was again recognised that this omission could simply be the result of a listener forgetting to include the descriptor for a source which was actually localisable. Panelists decided that the process of validating the graphical language would resolve this issue; since asking listeners to repeatedly describe their experiences using U-GAL descriptors would highlight whether an omission was erroneous. Listeners
decided that the requirement for different sized graphical descriptors could also be assessed during the language validation process.

**Representing audio images**

Panelists were required to reach a consensus on how best to represent the individual ‘audio images’ (‘separate sources’, ‘voices’ or ‘separate sounds’ as they had become known during the process of consolidation) within an auditory event. The process was not straightforward as more than one audio image could exist within a vehicle (when, for example a gap existed between individual instruments in an ensemble) and each could consist of more than one instrument with its own definable ‘centre of sound’ (when instruments from the same ensemble occupied the same location within the vehicle). From the outset it was decided that each instrument present within a reproduced stimulus should have its own descriptive letter. This letter could be positioned within the ‘size of sound’ descriptor representing the audio image, to indicate the instrument or instruments being represented at that location. Following further debate and exercises in which panelists were asked to either depict a possible scenario or to interpret an existing graphical representation, it was decided that audio images could be described as illustrated in figures 3.7.2, 3.7.3 and 3.7.4.

*Figures 3.7.2 - 3.7.4 representing audio images*

- Figure 3.7.2 audio image with 2 FOCAL
- Figure 3.7.3 two audio images
- Figure 3.7.4 image with & without FOCAL

The three different scenarios illustrated in the above figures can be described as follows: Figure 3.7.2 describes a situation where two localisable (with ‘centre of sound’) sources are present within the same audio image; figure 3.7.3 illustrates a situation where one localisable source is located at two distinct locations within the vehicle and therefore requires description using two ‘size of sound’ descriptors; and figure 3.7.4 uses one ‘size of sound’ descriptor to represent a single audio image containing two instruments where only one of these sources (P) is localisable.
3.8 Discussion and conclusions

Even though all panels started with a disparate set of individual graphical and verbal descriptors, similarities did exist between panels when these descriptors were consolidated in the panel languages. Most panels chose to describe the ‘area’ or ‘size’, ‘filled’ or ‘covered’ by an ‘audio image’, ‘soundstage’ or ‘separate sound’, with one panel describing the ‘shape’ of each source in terms of its left-to-right and front-to-back extent. When panels deconstructed the trio ensemble and solo instrument programme material into an audio image (or audio images), the locations of the sources within each auditory event were used to define the boundaries of a single audio image and only sources present at the same location were said to occupy the same audio image. A single ‘size of sound’ or ‘area’ descriptor was typically used to represent each audio image and within this, panelists developed graphical descriptors to represent the ‘focal point’, ‘main part’, ‘centre of sound’ or ‘origin’ of each localisable source.

Panels were in less agreement when it came to representing other attributes of their experiences. Some panelists believed it necessary to describe the ‘height’ (the position of a source on the vertical axis through the vehicle) of a stimulus; although most acknowledged the difficulty in representing height using a two-dimensional plan response form. Other panels spent time discussing the concepts of an audio image ‘involving’ or ‘enveloping’ the listener, the relative loudness of different instruments within an ensemble, and the depth of a solo instrument source. Two further panels spent considerable time discussing the concept of ambience or ‘feeling of space’, but only one went on to describe this experience graphically as a ‘cloudy’ ill-defined shape which could exist alongside the more definitive audio image. The cloudy descriptor was one of the few descriptors included in the universal graphical language - although (once again) there was much debate as to the existence and possible benefits of enabling listeners to describe a ‘feeling of space’ alongside the reproduced audio event. That said, the similarities between all panels’ languages made the development of the single, inter-subjective universal set of graphical descriptors (with associated verbal terminology) more straightforward for most of the auditory spatial experiences which required description. Finally, it was decided that a language of four distinct graphical descriptors could be used by a listener to describe all spatial attributes experienced when listening to the selected programme material. These were: A cloudy ‘feeling of space’; rectangular ‘size of sound’; circular ‘centre of sound’ and the denotative letters for sources present within each individual ‘separate sound’ (see figure 3.7.1).

3.8.1 Conclusions

Results of previous graphical elicitation investigations highlighted ambiguities in listeners’ responses when communicating their experiences of the width and location of reproduced audio stimuli. Although ambiguities could be caused by differences in the experiences of the individual listeners, listeners choosing to tackle the descriptive task in different ways, or representing differences in their individual experiences over time, it was also possible that ambiguities were being caused by factors of
Developing U-GAL

Chapter 3 - Development

the investigation setting; such as the listener misunderstanding the written research instruction. As an elicited response could only be considered valid if this reflected the experience of the listener, it was important that the noted ambiguities were clarified. The clarification of listener's individual graphical descriptions identified that the research objective (briefly - to develop a descriptive graphical language which would enable the graphical description of a listener's individual auditory spatial experiences and the comprehension of these experiences by the researcher) had yet to be fulfilled, since listeners appeared to be describing alternative descriptors to those requested in the investigation instruction. By asking listeners to describe the width of individual instruments and trio ensembles as perceived, it emerged that listeners were being asked to represent an attribute not readily experienced or easily represented using a graphical medium. Consequently listeners either attached this uni-dimensional attribute to their multi-dimensional experience with a degree of ambiguity; described the width attribute rather than their auditory experience or, conversely, described their experience rather than this attribute. In each situation there existed the possibility that the listeners' experiences would be misunderstood by the researcher, thus preventing the research objective from being fulfilled.

Via the process of asking listeners to revisit their individual graphical descriptions and verbal terminology within a group setting, a single - mutually acceptable - language of graphical descriptors and associated verbal terminology was developed. It was envisaged that this inter-subjective descriptive language would:

- Further both the researcher’s and listeners’ knowledge of the experiences which could be (and had previously been) represented using graphical and verbal descriptive languages; in this way improving the validity of the researcher’s existing and future accounts of listeners’ experiences.
- Provide the researcher with a set of verbal descriptors appropriate for; the definition of research questions; the provision of information to listeners and; the composition of a descriptive account representative of the listeners' auditory spatial experiences. By enabling the researcher to communicate in the same language as the listener, it was believed that the listeners' comprehension of any provided verbal language would be improved. Consequently, it was envisaged that listeners would be able to correlate the researcher's instructions with their (graphically describable) experiences. And, any subsequent explication by the researcher would be based on the listeners' experiences rather than on an inappropriately worded instruction.
- Minimise the problem of misunderstanding which could still result from listeners using their own individual graphical descriptions to represent their own individual experiences even when written instructions were appropriate. With the use of individual descriptors it would always be possible for the researcher to misconstrue the graphical descriptors used by the listeners and arrive at a descriptive account unrepresentative of their experiences. In developing U-GAL it was envisaged that the researcher would be able to base their descriptive account of the listeners' experiences on the listeners’ agreed use of a set of graphical descriptors rather than having to identify how each listener was using their own unique descriptors to represent their own experiences.
Thus the development of a universal graphical language (U-GAL) and set of associated verbal descriptors was believed to have minimised the sources of ambiguity previously identified within the communication process from researcher-to-listener (by improving the verbal terminology of the investigation question) and listener-to-researcher (via the development of a universal set of graphical descriptors). Accordingly the development of U-GAL was thought to have made further steps towards satisfying the research objective.

Figure 3.8.1 illustrates this progression towards an effective descriptive process using the graphical language development model\textsuperscript{15}. If, as suggested, the written instruction (Q) provided by the researcher reflects the experience of the listener, a listener will no longer have to choose whether to represent their experiences (Exp) or respond to the question of the researcher - thus minimising this previous source of ambiguity and improving the likelihood that the researcher will comprehend (CR) the listeners' experiences when represented graphically. However, there exists one foreseeable problem for the U-GAL: that listeners may not be able to represent their individual experiences using the developed language. In this situation, the researcher will still run the risk of miscomprehending the listeners' experiences (CM\textsuperscript{3}). Thus, as eluded to in the previous section, although an initial inter-subjective language has been developed by the listeners for the description of their auditory spatial experiences, a period of language evaluation is required before the effectiveness of this language can be established.

\textsuperscript{15} A key for the descriptive process model is provided in appendix I
3.9 Chapter summary

Chapter 3 has provided details of two studies which have aimed to further the development of the descriptive graphical language. Firstly, in the opening sections of the chapter (sections 3.1 – 3.5), a summary has been provided of an investigation which sought to clarify ambiguities in listeners’ individual graphical descriptions. The second part of the chapter has detailed the development of a universal graphical language (sections 3.6 – 3.8) for the representation of individual listeners’ auditory spatial experiences.

The clarification investigation determined that listeners were graphically representing multi-dimensional auditory spatial experiences and using multi-dimensional verbal terminology (such as area) to describe their representations. Accordingly, the uni-dimensional descriptors of instrument and ensemble width (used by the researcher in written instructions to participants in earlier GAL investigations) were identified as a likely cause of the ambiguities in listeners’ graphical responses.

In section 3.5 it was determined that, even if the terminology used in the written instruction was relevant to listeners’ auditory spatial experiences, the use of individual graphical languages could still prove problematic, because the researcher would not necessarily understand this graphical terminology in the same way as the listener intended.

The second part of the chapter detailed how a process of inter-subjective small group discussions enabled listeners to develop a mutual language of graphical descriptors for the description of their individual auditory spatial experiences. The development of this universal graphical language (U-GAL) required all listeners to agree on the mutual descriptors and the individual experiences that would be represented by these common graphical terms. A language of verbal terminology was also developed by the listeners to enable the issuing of effective written instruction by the researcher.

Following the development of the U-GAL (and the associated verbal descriptors) chapter 3 concluded by stating how the descriptive process had improved from earlier investigations. For, rather than having to identify how each listener used their own descriptors to represent their own experiences, the researcher could now base their understanding on an agreed application of graphical descriptor to experience, minimising ambiguities in their comprehension of listeners’ auditory spatial experiences. Furthermore, the development of the associated verbal descriptors increased the likelihood of the researcher using appropriate terminology in written instruction for future GAL investigations.

However, potential improvements to the descriptive process (brought about by the development of U-GAL) were currently unproven. There was still the possibility that a listener would not be able to associate the written instruction with their experiences. Moreover, with the development of an inter-subjective language, one prospect was that listeners would not be able to represent their individual experiences using this universal language. Consequently, the next stage in the development of the descriptive graphical language was to return the U-GAL to the listeners for evaluation.
CHAPTER 4

Evaluating U-GAL
4.0 Chapter overview

Chapter 4 recounts the process of evaluating the universal graphical language (U-GAL).

As detailed in chapter 3, the U-GAL had been developed by listeners for structuring and representing their auditory spatial experiences in a manner which could be comprehended by the researcher. Chapter 4 is specifically concerned with evaluating the effectiveness of this universal language by providing answers to the following questions:

1) Will individual listeners be able to use the U-GAL in place of their own individual language to structure and represent their individual auditory spatial experiences?
2) Will the researcher be able to comprehend the listeners' auditory spatial experiences when described in the universal language?

Details of the investigation setting for the evaluation of U-GAL are provided in section 4.2. Included in this section is information about the listeners involved in the study, the programme material and the investigation procedure.

To establish the suitability of the universal graphical language, sections 4.3 and 4.4 offer an analysis of the graphical depictions elicited from listeners during the evaluation investigation.

In section 4.3, an examination of listeners' individual descriptions identifies how the universal language was being used by the investigation participants. This analysis establishes potential sources of ambiguity for the researcher tasked with comprehending a listener's auditory spatial experiences through their graphical descriptions.

The statistical and graphical analyses in section 4.4 determine the listeners' ability to use the universal descriptors to represent differences in their individual experiences. Once again, ambiguities in listeners' responses are identified and, where possible, clarified by the researcher.

Discoveries from both sections combine to provide answers to the two questions introduced above, and, correspondingly, ascertain whether the development of the U-GAL has enabled the research objective to be attained.

A brief summary of the findings from chapter 4 and conclusions based on these findings are offered in sections 4.5 and 4.6.
4.1 Introduction

In developing the universal graphical language (U-GAL) it was assumed that improvements (with respect to previous stages in the language's development) would be made to the descriptive process. Specifically, it was believed that the development of U-GAL would enable the researcher to comprehend listeners' auditory spatial experiences when represented using this common medium. The reasoning behind this assumption was that the researcher would no longer have to understand how each listener was using their own unique descriptors to represent their own experiences. Instead, with all listeners using the U-GAL, it was believed that the task for the researcher would be less ambiguous. It was envisaged that, as each descriptor had been developed to represent a specific spatial attribute of an audio reproduction (as experienced), the researcher would be able to communicate the listeners' experiences by describing their (the listeners') use of the universal graphical descriptors. Furthermore, in developing an accompanying set of verbal descriptors, it was anticipated that the researcher would be able to communicate in the same language as the listener; one reflective of the listeners' auditory spatial experiences and the graphical descriptors created to represent these experiences.

However, even though the language development process had provided listeners (and the researcher) with an opportunity for obtaining a greater understanding of their auditory spatial experiences via the inter-subjective description of these experiences and had underlined similarities between listeners' experiences\(^1\), the development of U-GAL brought with it one obvious question: Would an individual listener who had previously represented their own experiences using their own graphical descriptors be able to structure and represent their own experiences using the inter-subjective descriptive language? To answer this question, and to identify whether descriptions using the universal language did indeed provide the researcher with a means of unambiguously comprehending a listener's experiences, U-GAL required a period of evaluation.

**Defining the terms of the evaluation**

When Berg (2002, p31) validated a selection of verbal attribute scales for the evaluation of spatial quality in reproduced audio, he hypothesised that the scales would be relevant (that is to say they would have "sufficient common meaning" for a group of listeners) if the listening group was able to use the scales to differentiate between "some or all of the stimuli in the experiment at a significant level". In common with Berg's hypothesis, a U-GAL user would be required to differentiate between their experiences of different audio stimuli using the inter-subjective descriptors in order that the language could be deemed relevant. However, in contrast with the aims of Berg and proponents of descriptive analysis techniques\(^2\) U-GAL's evaluation concentrated on the appropriateness of the mutual language for representing individual listeners' auditory spatial experiences rather than the experiences

---

\(^1\) See figure 3.6.1 in Chapter 3 - the similarities between the five graphical languages developed by the different panels.

\(^2\) See Bech (1996), Civille and Lawless (1986), Hootman (1992), Koivunen and Zacharov (2001), Møllgaard et al. (1999), and Zacharov and Koivunen (2001a, 2001b)
of a group. The reason for this departure was that although all listeners may describe a common experience when presented with the same stimulus (especially following a period of inter-subjective description and discussion), the situation as experienced could still be different for different listeners. Thus evaluation of the U-GAL reflected this possibility and a hypothesis constructed as follows:

If a language of universal descriptors has sufficient meaning for a listener - in other words if U-GAL is a suitable medium through which a listener may communicate their auditory spatial experiences - the listener will be able to use the language to represent differences in reproduced audio stimuli where differences in experiences exist.

At the most rudimentary of levels, it was believed that U-GAL could be considered a suitable communication medium for listeners if it was established (through statistical analysis) that they could use the graphical descriptors to represent differences in the spatial characteristics of selected stimuli. However, as the listeners’ graphical depictions were required to represent their individual experiences of stimuli (the stimuli as it appeared to them) results of a statistical analysis, which would be based on stimuli as experienced by the researcher, were considered limited in their usefulness for determining the suitability of the U-GAL. Since listeners were not expected to describe spatial differences which were not experienced, results of statistical analysis had the potential for being misleading where non-significant results (those where no appreciable spatial differences were being described by a listener) were reported. Further evaluation of non-significant responses was therefore required to establish whether these results were due to the unsuitability of the descriptive language, some other factor of the investigation setting, or the listener not experiencing any differences.

Psychological literature (see Coolican 1996) cites reliability as an appropriate indicator of the robustness or stability of a measure; the measure, in this instance, being a universal graphical descriptor. Thus, the consistency of individual listeners when using the descriptive language was expected to provide a further indication of U-GAL’s suitability for the task of describing their experiences. Accordingly, the language could be considered reliable in situations where listeners used the universal graphical descriptors consistently to repeatedly describe a stimulus as experienced. However, this statement was, once again, made with a proviso: Since a listener could have a different experience of the same stimulus over time, they may consequently (and quite validly) change their graphical representation to correspond with this differing experience. Thus a situation was possible where descriptive inconsistency reflected the experiences of the individual and not the unreliability of the language. Provision therefore needed to be made for any consistency analysis to acknowledge the influence of the individual experience. It was decided that an examination - a graphical analysis - of each listener’s responses would provide the researcher with an indication of listeners’ consistency and a means of detecting where, and describing why, inconsistencies were occurring.

Two complimentary methods were therefore proposed to evaluate listeners’ ability to use U-GAL descriptors. These methods were: i) a statistical analysis of the spatial differences represented by the

---

3 The introduction to Chapter 3 outlines a list of the common causes of ambiguity in subjective sensory evaluation.
individual listeners and ii) a graphical analysis of listeners’ individual responses - to establish reasons for any non-significant differences and to provide an indication of individual listener consistency. By evaluating the suitability of U-GAL descriptors for representing listeners’ experiences, it was possible to appraise the universal language for its ability to progress from experience to representation in the descriptive process model\(^4\) and to provide an answer to the first question identified in the overview to this chapter; whether individual listeners would be able to use the U-GAL in place of their own individual language to structure and represent their individual auditory spatial experiences.

However, as U-GAL was primarily developed to improve on the second part in the descriptive process (getting from listeners’ graphical representations to the researcher’s comprehension) it was necessary to evaluate U-GAL for its descriptive suitability in this respect, and to establish whether the researcher would be able to comprehend the listeners’ auditory spatial experiences when described using U-GAL.

Although the potential for misinterpretation by the researcher was assumed to have been minimised by the development and use of the universal graphical descriptors, there still existed the possibility that a listener would not use the developed descriptors (even if considered appropriate for the description of their experiences) to represent their experiences as had been agreed during U-GAL’s development. In such circumstances, the researcher’s comprehension of each listener’s experiences would not be appropriate (representative of the listeners’ experience) as this would be based on the agreed use rather than any alternative use of a descriptor.

A second supposition that required evaluation was that listeners involved in the U-GAL’s development would be able to understand what was being asked of them if their universal verbal language was used in the provision of any written instruction by the researcher. Previous studies had indicated that a listener unable to relate provided information back to their experiences would be more likely to mould any unsuitable question to fit their experiences rather than respond to this written instruction. This action could result in an increased level of ambiguity; with the listener representing something other than their experiences (a problem for both stages in the descriptive process) or representing their experiences and not following the investigation instructions; a specific problem for the researcher.

Consequently the assumption that listeners would be able to respond to the written instruction provided by the researcher using U-GAL descriptors as agreed needed evaluation before the research objective could be met. It was decided that these suppositions could best be assessed by looking at each listener’s individual use of the inter-subjective graphical descriptors. Only through this second graphical analysis would it be possible to identify whether listeners had understood the terminology provided and whether the inter-subjective descriptors were being used as intended by the individual listeners.

The processes and results obtained from the ensuing evaluation are discussed in the remainder of this chapter with supporting material (visual plots and graphs) included in the appendices of Appendix 4.

\(^4\) See, for example, Chapter 1, figure 1.4.1

N Ford Doctoral Thesis 150
4.2 Evaluation process and investigation requirements

The evaluation of the universal graphical language (U-GAL) observed procedures similar to those employed when eliciting graphical descriptors from individual listeners within the clarification investigation\(^5\). One notable difference was that in this evaluation study, participants were asked to describe their experiences using the inter-subjective U-GAL descriptors rather than their own individual language. As with the clarification study, the listener sat in the (right-hand) driver's seat of the same motionless vehicle. The vehicle was again equipped with a multichannel audio system capable of up-mixing the two-channel stereo stimuli into seven loudspeakers.

4.2.1 Programme material

Eight items of stimuli were used in the evaluation. These musical extracts originated from the Archimedes CD of mono anechoic recordings and were a subset of those previously used in the clarification study. Chapter 3 (section 3.2.1) and table 4.2.1 contain information about the stimuli.

<table>
<thead>
<tr>
<th>Name</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Percussion Left (Perc L)</td>
</tr>
<tr>
<td>2</td>
<td>Percussion Central (Perc C)</td>
</tr>
<tr>
<td>3</td>
<td>Cello Left (Cello L)</td>
</tr>
<tr>
<td>4</td>
<td>Cello Central (Cello C)</td>
</tr>
<tr>
<td>5</td>
<td>Cello Percussion Voice: Narrow (CPVN)</td>
</tr>
<tr>
<td>6</td>
<td>Voice Cello Percussion: Wide (VCPW)</td>
</tr>
<tr>
<td>7</td>
<td>Voice Cello Percussion: Left (VCPL)</td>
</tr>
<tr>
<td>8</td>
<td>Cello Percussion Voice: Right (CPVR)</td>
</tr>
</tbody>
</table>

The stimuli had been chosen and processed\(^6\) to emphasise different spatial characteristics. Although individual instrument sources within the four different trio ensembles (stimuli 5-8) were not manipulated in terms of their individual size, the total area covered by each ensemble was varied. For example, the area covered by the wide trio ensemble VCPW (stimulus number 6) was recorded to contrast with that occupied by the narrow ensemble CPVN (stimulus 5), with individual instruments amplitude panned to the far left, centre and far right of the reproduced stimulus for the VCPW ensemble and clustered around the centre for the CPVN extract. The four solo instrument sources did not explicitly change their size but (as with the clarification investigation) these were amplitude panned to different positions; far left and central. The decision to use a selection of the stimuli from the clarification study avoided introducing another variable into the language evaluation process and also meant that an informal comparison was possible between responses provided in listeners' individual graphical languages and using the universal descriptors\(^7\).

---

\(^5\) See chapter 3 for full details of this study

\(^6\) Material was prepared by Russell Mason at the University of Surrey

\(^7\) It is accepted that by limiting the programme material and using the same stimuli throughout the development of U-GAL, the validity of the language in other contexts (where different stimuli is used) may be somewhat restricted.
4.2.2 Investigation listeners

Twenty-two listeners evaluated the suitability of U-GAL for describing listeners' individual auditory spatial experiences. Eleven of these listeners had participated in the development of the universal descriptor set by attending initial or initial and final stage discussion sessions. These same listeners were also adept at representing their experiences using their individual graphical descriptors having participated in previous I-GAL studies. The remaining listeners had no experience of using graphical descriptors (either individual or universal) to represent their experiences. By including listeners in the study with no prior experience of using graphical descriptors, a measure of the external validity of the inter-subjective language could be established: In other words, it would be possible to identify if U-GAL could be used by listeners other than those involved in its development.

As existing GAL users were not necessarily trained listeners and conversely listeners new to U-GAL may have had previous training in listening or possessed relevant listening expertise (for example as practical musicians), different levels of expertise were possible within each listening group. Since the experience of listeners is acknowledged as an influential factor in listening investigations each listener's level of expertise was identified using a simple questionnaire. Listeners were asked if they had received any training in how to listen to reproduced audio, whether they were able to play a musical instrument after following a formal training program or self-instruction, or if they possessed no such experience. Listeners were initially divided according to four groups - classifying that they had received: (i) specific training in how to listen to and evaluate reproduced audio; (ii) formal training in playing a musical instrument; (iii) no formal training, but a self-taught musician, or; (iv) no previous training in either evaluating reproduced audio or playing a musical instrument. However, listeners were eventually divided into two broad groups: (i) listeners with prior training of evaluating reproduced audio and (or) experience of playing a musical instrument and (ii) those who had never played a musical instrument or received any training in listening. The division of listeners into two groups was to ensure (for the purposes of statistical analysis) that a similar number of listeners was present in both groups. An indication of how listening groups were divided according to experience and the listeners' prior acquaintance with a graphical descriptive language is outlined in table 4.2.2.

Table 4.2.2 Listening group composition

<table>
<thead>
<tr>
<th>Listener Characteristics</th>
<th>Number of listeners</th>
</tr>
</thead>
<tbody>
<tr>
<td>Existing listeners with previous listening training or who play(ed) a musical instrument</td>
<td>6</td>
</tr>
<tr>
<td>Existing listeners with no relevant listening training or musical experience</td>
<td>5</td>
</tr>
<tr>
<td>New listeners with previous listening test training or who play(ed) a musical instrument</td>
<td>5</td>
</tr>
<tr>
<td>New listeners with no relevant listening training or musical experience</td>
<td>6</td>
</tr>
</tbody>
</table>

Although it is acknowledged that experience obtained through playing a musical instrument is different from the specific experience obtained when a listener is trained to evaluate audio products, an ability to listen is fundamental for the successful completion of either task. Furthermore, Bech (1989) includes

---

*Authors who have written about the Influence of listener experience include Kirk (1956), Gabrielsson (1974, 1975), Bech (1989, 1992) and Toole & Olive (1994)*
both experiences as factors which could potentially influence a listener’s ability to assess reproduced sound. Consequently, the division of listeners according to these two broad groups was deemed suitable for the task of identifying whether this listening experience would influence a listener’s responses.

4.2.3 Evaluation process

During the investigation, listeners were asked to provide graphical representations of their experiences using the developed U-GAL descriptors and a computerised response gathering form. The response form was presented to listeners on a touch-screen which listeners could hold flat on their laps within the vehicle (as they would a piece of paper) and use a pen-like pointer or their finger to operate. Prior to starting the investigation all listeners undertook a tutorial which familiarised them with the computerised response method and the layout of the U-GAL screen. No specific information about the actual investigation was provided at this stage. Tutorial extracts are included on page 1 of Appendix 4.A. (figures 4.A.1-4.A.6).

After completing the tutorial and prior to starting the investigation, listeners were required to read an instruction sheet detailing how each of the provided descriptors should be used to represent their experiences of what they were about to hear. The instruction sheet used only terminology developed alongside the universal graphical descriptors in order that the verbal language corresponded with the graphical task and the possible experiences of the listener. A copy of the instruction sheet is included in Appendix 4.A.2.

Three trial runs were provided for each listener, in order that they could develop their representation style using the graphical interface and touch-screen before commencing the actual investigation. Trial runs were the same for each listener. The investigation itself consisted of 24 runs - one for each stimulus item, with each appearing three times to enable the evaluation of individual listener consistency. On commencing the actual investigation each listener was given a unique user ID which corresponded with one of six different randomised run orders. Each run order was linked with a visual grid of 24 boxes (see figure 4.2.1) with each box linked to one musical extract. By clicking on a grid box, listeners could access a U-GAL response page (figure 4.2.2) where they could use the provided graphical descriptors to represent the item of programme material being played. After describing the stimulus, listeners clicked on save and continue on the response page, responses were saved, the programme material stopped playing and the listener was taken back to the grid. Listeners could choose in which order to complete the grid and could return to a particular stimulus - to edit or review their representation - by clicking on a completed box which was denoted by a tick (see figure 4.2.3). To avoid encouraging listeners to match the stimuli within the grid and editing their representations based on this information they were not informed that repeats would be present. The task for the listener was completed when all 24 boxes contained ticks.

All programming for the computerised U-GAL response form (U-GALUI) was undertaken by Toby Newman of Harman/Becker Automotive, Bridgend.
Data collected from all listeners existed in the form of a computerised database. Information in the database could be recalled by run number and included (for each run) a record of the listener’s unique reference number, the programme material reproduced, which U-GAL descriptors had been used to describe the experience of the listener and the co-ordinates and two-dimensional size of these descriptors. With the data in place, analyses could be conducted to evaluate the universal language.

A statistical analysis of the spatial differences represented by each listener when auditioning different stimuli was conducted to provide an indication of whether the universal language could be used by an individual listener to describe their experiences. For U-GAL to be deemed appropriate for individual listener use, a listener was required to demonstrate sufficient consistency in their graphical responses to differentiate between the different stimuli when differences in their spatial attributes were experienced. Thus language evaluation also involved a graphical analysis of each listener’s descriptions to identify plausible reasons for inconsistencies in a listener’s responses when these occurred (in essence, why the same descriptor was being used differently over time) and providing, where possible, explanations when differences in the stimuli were not being represented by the individual listeners.
A second graphical analysis also provided evidence about how each listener was using the descriptive language; whether U-GAL was being used as agreed, or if the listener were appearing to mould the language to represent alternative or additional experiences. Identifying how listeners were using the developed descriptors was essential for ensuring the research objective was met. If a graphical descriptor was being employed by a listener in a manner contrary to that intended, it could firstly indicate that the listener was unable to use the inter-subjective language to represent their auditory spatial experiences, and secondly, it could have repercussions for the researcher when attempting to comprehend the listener’s experiences through the listener’s descriptions. By creating visual plots which overlaid all three of a listener’s responses to the same stimulus, not only could the level of consistency achieved by a listener be explored, but it was also possible to obtain an indication of how each listener was choosing to use the different U-GAL descriptors to represent their individual experiences.
4.3 Establishing how listeners used U-GAL : The graphical analysis of listeners’ descriptions

It is advantageous at this time to revisit the descriptors developed by the listeners, to understand their agreed use for representing a listener’s experiences, in order to establish how listeners should have been using the inter-subjective graphical descriptors to represent their experiences.

Rather than divide a reproduced stimulus into individual sources or complete ensembles and describe the spatial properties of each of these distinct auditory events, listeners developed U-GAL to describe the spatial attributes of ‘audio images’ or ‘separate sounds’ as experienced. A separate sound (also known as a ‘sound’ in this evaluation) was defined as being either a solo instrument, or group of instruments from an ensemble stimulus which occupied the same region in space at the same time. The spatial characteristics of each sound could then be described by the graphical descriptors outlined below.

Participating panelists\(^10\) agreed that the area covered by a sound could be represented using the (round-cornered) rectangular ‘size of sound’ descriptor, ‘\[
\begin{array}{c}
\text{\hspace{1cm}} \\
\end{array}
\]’ identified in this evaluation using the term AREA. To represent listeners’ experiences of the size of the sound, the AREA was developed to change size in two directions; front-to-back and left-to-right.

A circular descriptor ‘\[
\begin{array}{c}
\text{\hspace{1cm}} \\
\end{array}
\]’ (referred to in this language evaluation as FOCAL) was created to represent the ‘centre of sound’; a localisable centre or core which could exist within the AREA to indicate the main part of a reproduced stimulus. The number of FOCALs which could be used by a listener in the same sound was not restricted. The FOCAL descriptor could not change size.

If used as agreed, a third shape ‘\[
\begin{array}{c}
\text{\hspace{1cm}} \\
\end{array}
\]’ was expected to be positioned independently of, or surrounding the AREA to represent a ‘feeling of space’ within the vehicle; essentially whether the listener felt themselves to be listening in a small or large environment. This descriptor, known henceforth by the initials FOS, was developed to change size in two directions; front-to-back and left-to-right.

Graphical descriptors were introduced to naïve listeners\(^11\) (and re-introduced to the existing listeners) in the Instructions for Listeners Using U-GAL sheet, provided for reference in Appendix 4.A.2. If listeners used the developed descriptors as agreed, representations similar to those illustrated in figures 4.3.1 and 4.3.2 were expected\(^12\). If used as agreed, figure 4.3.1 could be interpreted as representing a focused cello at the centre of the vehicle with a ‘feeling of space’ covering the driver and passenger.

\(^{10}\) 20 panelists were involved in the development of U-GAL and associated verbal descriptors. Eleven of these panelists participated in the evaluation stage.

\(^{11}\) Naïve listeners were categorised as those who had not participated in the development of U-GAL.

\(^{12}\) Listeners were not shown these exemplary depictions prior to participating in the language evaluation. However, they had previously participated in the U-GALUI tutorial. The tutorial demonstrated how listeners could manipulate the various descriptors. (see appendix 4.A.1 for extracts from the tutorial)
seats. Figure 4.3.2 could be interpreted as representing an unfocused cello source, again at the centre of
the vehicle with the ‘feeling of space’ surrounding this instrument.

![Focused cello with a separate feeling of space](image1)

![Unfocused cello with a surrounding feeling of space](image2)

When not presented for immediate reference alongside the following text, graphical plots are contained
in appendices 4.B and 4.C. When visual plots for each listener were examined, it was identified that
U-GAL descriptors were being used as agreed by some listeners.

Listener 2 (page 4.B.1, plots 1-8), Listener 10 (page 4.B.1, plots 9-16), Listener 1 (page 4.B.2, plots 1-8), Listener
22 (page 4.B.2, plots 9-16)

The graphical descriptions of two existing listeners (listeners 2 and 10) and responses from two
listeners with no previous experience of using a graphical language or developing the U-GAL (listeners
1 and 22) exemplified that, regardless of their experience, U-GAL descriptors could be used as
expected. For these four listeners the FOCAL descriptor (when used) represented a localisable centre
within an AREA, and an FOS was used to surround this AREA on those occasions when a feeling of
space was experienced. For these exemplary listeners, a letter identifying the instrument (a ‘P’, ‘V’ or
‘C’) was positioned within a FOCAL where this existed. Figures 4.3.3 and 4.3.4 illustrate how a new
listener (L1) used the graphical descriptors as agreed, whilst figures 4.3.5 and 4.3.6 exemplify how the
listener 2 was an existing listener with many years experience of engineering sound systems within vehicles and
listener 10 was an existing listener, untrained musician and loudspeaker engineer

---

13 The ten pages of plots in Appendix 4.B are named 4.B.1 to 4.B.10 and contain responses from individual listeners. On each page, 16 plots are presented: Eight each for two different listeners. Each of the plots is constructed from all the descriptors used by an individual listener when repeatedly describing the same item of programme material. The plots on each page are numbered from one to 16 with numbers one-to-eight referring to the descriptions of one listener and nine-to-16 to those of the second listener on the page. Consequently, when graphical plots are referred to from the main text, (typically in small italics at the start of a sub-section) both a page reference (for example 4.B.1 or 4.B.2), and a plot number (1-16) will be displayed. A similar lexicon will be used when referring to plots from appendix 4.C. In this appendix, plots are separated by U-GAL descriptor use rather than by listener, with individual plots consisting of responses for all listeners who used the same descriptor to describe the same stimulus. Both appendix 4.B and 4.C have an introduction containing further information about their layout and contents.

14 Listener 2 was an existing listener with many years experience of engineering sound systems within vehicles and listener 10 was an existing listener, untrained musician and loudspeaker engineer.
existing listener, (L2) described the same stimuli; the central cello source and the ensemble of 'cello, voice, percussion right' (CPVR).

4.3.1 Use of U-GAL letters

Not all listeners employed U-GAL descriptors as agreed. Several listeners separated the FOCAL descriptor from the identification letters. Exemplary responses for two such listeners, listener 3 (L3) and listener 6 (L6)\(^{15}\), are illustrated in figures 4.3.7 and 4.3.8 with further examples provided in appendix 4.B. The separation of the FOCAL and letter descriptors did not in itself represent a departure from the agreed use of U-GAL: Listeners were instructed to use FOCAL to indicate the presence of a localisable core within a region and to use letters to identify the instrument being described. No specific instruction was offered as to where identification letters should be positioned and although letters had been positioned within a FOCAL descriptor during the tutorial, this positioning was not essential. However, as letters were not expected to be positioned with any forethought, any consistency when positioning letters independently of the FOCAL descriptor was unexpected and did indicate a departure from U-GAL’s anticipated use.

\(^{15}\) Listener 3 had no listening experience outside of participating in GAL investigations, whilst listener 6 was new to both graphical elicitation and listening investigations
Listener 13 (page 4.B.3, plots 13-16), Listener 18 (page 4.B.5, plots 1-8), Listener 11 (page 4.B.6, plots 1-8). Consistency in letter placement did not only occur alongside empty FOCAL descriptors. For many listeners (including listener 13 (figure 4.3.9), listener 18 (figure 4.3.10) and listener 11) U-GAL letters appeared to replace the FOCAL descriptor, with letter placement mimicking the placement of FOCAL in other listeners.

The use of FOCAL and letter descriptors

This observation could be further supported by an examination of the graphical plots in Appendix 4.C. Similarities between plots 9-12 on page 4.C.1 (displaying the letters used by listeners without FOCAL descriptors) and plots 1-8 on page 4.C.2 (illustrating the use of FOCAL descriptors to represent the same stimuli) suggested that identification letters, rather than FOCAL descriptors, may have been used by some listeners to represent a localisable core or focal point within the region covered by a ‘separate sound’. An illustration of the similarities in the use of FOCAL and letter only descriptors is provided in figures 4.3.11 & 4.3.12.

Further support for the theory that some listeners used letters in the place of the FOCAL descriptors was obtained via a comparison of the 16 plots on page 4.C.1. In the first group of plots (Appendix 4.C, page 4.C.1, plots 1-8) plotted letters were used alongside a FOCAL descriptor to represent the different items of programme material. A comparison could be made between these eight plots and those where letters were used without a FOCAL descriptor to describe the same stimuli (plots 9-16 from the same page). If listeners were solely using letters to identify the instrument represented within a particular AREA, the placement of letters in plots 9-16 would be expected to be more random than in plots 1-8. However, for many stimuli, the letter locations in the two sets of plots were similar, as exemplified by figures 4.3.13 and 4.3.14; responses for all listeners who used letters within FOCAL descriptors to represent the ‘cello, percussion, voice right’ (CPVR) ensemble (figure 4.3.13) and those who used letters without FOCAL descriptors to represent the same stimulus (figure 4.3.14).

Rather than indicating an alternative use for the letter descriptor, similarities in letter positioning may have been due, in part, to the size of the AREA being described: If an instrument or ensemble without a localisable centre or core was deemed to occupy a small region within the vehicle, little difference
would exist between the positioning of a letter within this small AREA and a similarly sized AREA with a FOCAL.

Figure 4.3.15 L8 (CPVN) inconsistency using letters
Figure 4.3.16 L9 (Perc L) use of FOS without AREA
Figure 4.3.17 L11 (Cello L) use of FOS without AREA

Listener 8 (page 4.B.8, plots 10, 13, 15)
It should also be noted that not all listeners used identification letters in a structured way. Figure 4.3.15 shows how listener 8, a new listener and trained musician, demonstrated little consistency when positioning letters. However, listener 8 also tended to represent experiences using the FOS descriptor without FOCAL or AREA. This in itself demonstrated a departure from the anticipated use of U-GAL descriptors and required further examination.

4.3.2 Describing the region a sound covers
Although the FOS descriptor was used by some listeners alongside an AREA to represent a feeling of space (see, for example, earlier figures; 4.3.3, 4.3.6 and 4.3.7), it was not used in this way by all listeners. In addition to listener 8, many listeners chose to use a FOS without employing an AREA descriptor. Listener 916 chose to represent their experience of the percussion left stimulus using a FOCAL surrounded by a FOS descriptor in two of the three repeated presentations of this stimulus (an occurrence captured in figure 4.3.16), whilst listener 1117 used the FOS descriptor without an AREA or FOCAL to surround letters when representing all sources (for example, figure 4.3.17).

Another existing listener and participant in the final U-GAL development sessions chose to use a FOS surrounding either a FOCAL descriptor or a letter when describing the ensemble stimuli (see figure 4.3.18). Within these ensembles, listener 15 represented the voice towards the rear of the vehicle inside a FOS, whilst other instruments were located at the front of the vehicle within an AREA. In previous GAL investigations, listener 15 had represented entire ensembles using circular descriptors (figure 4.3.19 and appendix 3.B, plots 3.B.56 - 3.B.59). It therefore appeared likely that the development and use of two distinct descriptors (FOS and AREA) had enabled this listener to structure their experiences by differentiating between the spatial characteristics of the ensemble instruments. Similarly, since

---

16 An experienced existing listener
17 An experienced existing listener and participant in the final language development meetings
listeners 9, 11 and 15 had, on occasions used both the FOS and AREA descriptors as agreed to represent their experiences, one possible conclusions was that - at least for these listeners - the FOS descriptor did not simply replace AREA as a means of describing the region covered by a sound. This possibility could be further substantiated by the fact that all three listeners were involved in the development of the inter-subjective descriptor set, with listeners 11 and 15 involved in the final development of U-GAL.

Thus it may be suggested that rather than mistakenly using the FOS descriptor to describe their experiences of the region covered by a 'separate sound', some listeners were intentionally developing the descriptive language to describe a difference in the experiences that could be represented by the FOS and AREA descriptors. Consequently a question needed to be answered, namely: How were the different descriptors being used? Essentially; how did spatial characteristics differ between stimuli when described by the FOS or AREA descriptor?

Differences between FOS and AREA regions
One explanation for listeners choosing to use a FOS rather than an AREA descriptor when representing a 'separate sound' was initially thought to be the size of the different descriptors. Because the FOS descriptor was designed to be able to cover a larger region of the vehicle than the AREA descriptor it was initially believed that listeners may have had to use the FOS descriptor in situations where an AREA was unable to meet their requirements. However an examination of visual plots indicated that the FOS was being used to describe regions which could equally have been represented using the AREA descriptor. Furthermore, when asked in a post-investigation questionnaire if the provided descriptors allowed for the representation of the 'size of the sound' experienced, listeners who used the FOS descriptor did not mention that the AREA descriptor restricted their description.

Because the FOS descriptor was developed to represent a feeling of space rather than a definitive body or area of sound, listeners may have been chosen this cloudy descriptor to represent stimuli perceived

---

18 Listeners who described small regions using the FOS descriptor included listener 9 (page 4.B.5, plots 9 and 11), listener 7 (page 4.B.7, plots 1, 3-5, 7-8), listener 11 (page 4.B.6, plots 1-8) and listener 15 (page, 4.B.6, plots 9-12).
19 Comments made by listeners in response to the post-investigation questionnaire are included in appendix 4.E.
as being enclosed by less rigid boundaries than those contained within a solid rectangular AREA; potentially using the FOS to represent instruments or ensembles experienced as being unfocused - themselves a ‘feeling of space’. To examine this theory further, density plots were created to help characterise the differences between regions which could be described by an AREA and those which required description using a FOS descriptor. An inspection of Appendix 4.C, page 4.C.3 enabled the comparison of solo instrument representations using the AREA descriptor (plots 1-4) and those represented using a FOS descriptor (plots 5-8), with figures 4.3.20 and 4.3.21 providing an immediate reference of how AREA and FOS descriptors were being used to describe the percussion left stimulus. From the density of the plots it was possible to determine that the majority of listeners had chosen to use the AREA descriptor to represent the region covered by each solo instrument. It was also possible to ascertain that the FOS descriptor was being used in a similar way to the AREA, as the highest density of FOS (in other words the darkest shading on each plot) corresponded with the region covered by the AREA descriptors. Although listeners did appear to be describing less focused regions using the FOS descriptor, similarities in descriptor use suggested that size alone was not responsible for a listener favouring one descriptor over the other, and that perceived differences in the structure (boundaries) of the sound were just as likely to effect the use of the alternative descriptor.

In Appendix 4.C, pages 4.C.4 and 4.C.5 contain plots depicting the individual instruments within each ensemble when described by the AREA descriptor (page 4.C.4) and FOS descriptor (page 4.C.5) without an AREA. These plots not only highlighted differences in the focus of the particular instruments but also illustrated that, in most instances, the AREA descriptor was used to represent more focused sources at the front of the vehicle whereas the FOS was used to represent instruments which extended towards and even past the listening position. A useful example was provided in the description of the voice in the *cello percussion voice narrow* ensemble. When the AREA descriptor was used to describe the voice (figure 4.3.22) the instrument was represented as a focused source at the front of the vehicle whereas the FOS description of the same instrument (figure 4.3.23) indicated that listeners had experienced a far less focused sound throughout the vehicle. Although this ensemble provided a useful distinction between AREA and FOS use, the voice in the remaining three ensembles was described similarly by both AREA and FOS descriptors. Thus, even though some listeners appeared to be using the FOS descriptor to represent a less focused or a more fuzzy feeling of sound,
others could have been mistakenly using a FOS descriptor instead of an AREA to represent the region covered by a sound.

**Investigating an additional function of the FOCAL descriptor**

The FOS descriptor was not alone in apparently having more than one purpose. In addition to its agreed role, representing a core or localisable centre within the region covered by a sound, the FOCAL descriptor was used independently of any surrounding AREA by a number of listeners.

*Listener 18 (page 4.B.5 plots 1-4), Listener 16 (page 4.B.7 plots 9-12), Listener 11 (page 4.B.6, plot 6)*

Figure 4.3.24 provides an example of how listener 18\(^{20}\) tended to use only the FOCAL descriptor when representing solo instruments. This method of representing the solo sources was also favoured by the new and inexperienced listener 16, and listener 9 (figure 4.3.25). The use of the FOCAL descriptor without a surrounding AREA was not restricted to describing solo instruments. Listener 9 (figure 4.3.26), listener 18 (figure 4.3.27) and listener 11 used a FOCAL descriptor to represent the percussion source within the ‘voice cello percussion wide’ (VCPW) ensemble.

Figures 4.3.24-4.3.29 Descriptions favouring the FOCAL descriptor & overlaid plots of FOCAL use, describing the central percussion without (figure 4.3.28) and with (figure 4.3.29) AREA descriptor

Figures 4.3.24 L18  Figure 4.3.25 L9  Figure 4.3.26 L9  Figure 4.3.27 L18  Figure 4.3.28 Figure 4.3.29

That these (mainly experienced) listeners chose to use the FOCAL descriptor without an AREA to describe stimuli is not inexplicable. One explanation is that focused sources were perceived as occupying only a localisable core where no region existed outside of this centre. Four of the plots on page 4.C.2 of Appendix 4.C (plots 1-4) show occurrences of FOCAL descriptor use without the presence of an AREA when representing the four solo instrument sources. The remaining plots (plots 5-8) on the same page reveal occasions when listeners used a FOCAL descriptor inside an AREA to describe the solo sources. When plots of the same central percussion stimulus item are compared in figures 4.3.28 and 4.3.29, there appear to be similarities between when a FOCAL was used independent of an AREA (figure 4.3.28) and when the descriptor was used within an AREA (figure 4.3.29). These similarities further support the suggestion that listeners who employed a FOCAL independently of an AREA descriptor may have chosen to do so in order to represent a localisable centre or core of a sound where no surrounding region existed.

\(^{20}\) a highly experienced listener, GAL user and automotive audio systems engineer
Representing the regions covered by different stimuli
Because of the different ways in which listeners were using U-GAL descriptors to represent the regions covered by the various stimuli, it was not only necessary to establish why listeners were using the graphical descriptors as they were, but also to ascertain whether differences between stimuli could still be described when listeners used their own interpretation of the language.

As the three instruments within the narrow ensemble (CPVN) had been amplitude panned to the centre of the stimulus and those in the wide (VCPW) ensemble panned to the far left and right as well as to the centre, it was anticipated that differences would be most likely to occur between the regions covered by these two ensembles: With instruments occupying distinct, well spaced ‘separate sounds’ in the wide ensemble but more likely occupying the same ‘separate sound’ in the narrow stimulus.

Figures 4.3.30-4.3.35 Differences in descriptions of the CPV Narrow and VCP Wide ensembles (listeners 2, 1 and 13)

Listener 10 (page 4.B.1 plots 13-14), Listener 15 (page 4.B.6 plot 13-14), Listener 7 (page 4.B.7 plots 5 - 6), Listener 16 (page 4.B.7 plots 13-14)

For listeners 2 (figures 4.3.30 and 4.3.31) and 10 a difference in the clustering of FOCAL descriptors was immediately apparent when the narrow (CPVN) and wide (VCPW) ensembles were described. Both listeners positioned descriptors representing the narrow ensemble (see figure 4.3.30 for an immediate reference) at the centre of the vehicle but spread these FOCAL descriptors out when representing the VCPW ensemble (figure 4.3.31). For listener 1, descriptors were once again more clustered when describing the CPVN ensemble (figure 4.3.32) than the wider ensemble (figure 4.3.33). However unlike for listeners 2 and 10, where FOCAL descriptors were moved around within a single AREA descriptor, listener 1 used different AREA descriptors and clustered these together at the centre of the vehicle when describing the narrower ensemble and spread the descriptors out over the front of the vehicle to describe the wider trio. Rather than changing the clustering of descriptors, listener 13 (figures 4.3.34 and 4.3.35) increased the size of the AREA, whilst listener 7 and listener 16 increased the size of the FOS descriptor when representing the VCPW ensemble. Similar changes in AREA and FOS descriptor size were represented by listener 11 (figures 4.3.36 and 4.3.37) and listener 15. Both listeners also changed the location of letter descriptors from clustered together when representing the narrow ensemble to further apart when representing VCPW.
Further differences in the clustering of instruments could be seen for listener 21 in figures 4.3.38 and 4.3.39 and listener 6. Again the narrow ensemble was clustered around the centre of the vehicle for both listeners with the wider ensemble spreading out over the front of the vehicle. For listener 18, a gap was present between the percussion instrument and the remainder of the ensemble when the listener described the wider of the two ensembles (as illustrated in the earlier figure, 4.3.27). The location of the percussion at the far right of the vehicle suggested that the VCPW ensemble had been split into two distinct ‘separate sounds’ by the listener, with the voice and cello occupying one region and the percussion occupying its own distinct ‘separate sound’ within the wide ensemble. Listener 20 also represented differences between the two ensembles by moving the wide ensemble instruments further to the left and right of the vehicle and increasing the gap between these instruments. The difference between ensembles is illustrated in figures 4.3.40 and 4.3.41.

Thus, an examination of graphical plots for individual listeners indicated that, although not all listeners were using the language as agreed, differences were being represented between the region covered by the narrow ensemble and that covered by the wider ensemble. Differences were mainly visible in the clustering of instruments, with FOCAL descriptors spread over more of the vehicle for the wider ensemble than for the narrow ensemble. Occasional differences were visible between the size of the AREA or FOS descriptors used to represent the two ensembles. Size differences tended to occur when listeners used one descriptor to represent a single ensemble rather than dividing the stimulus into ‘separate sounds’ and representing each instrument separately within its own AREA or FOS. In situations where listeners had chosen to represent individual instruments using separate AREA or FOS descriptors, these were more likely to overlap when representing the narrow ensemble. Further conclusions resulting from the graphical analysis of visual plots are presented in the following section.
4.3.3 Graphical analysis conclusions

The primary concern of the graphical analysis was to establish how individual listeners were using U-GAL’s graphical descriptors. It was identified through this analysis that, although some listeners followed the instructions provided for them and used the developed language as agreed, other listeners appeared to derive a more individual approach when applying the descriptors to their experiences.

For many listeners it emerged that letter descriptors were taking on an alternative role to that agreed. In this alternative role, it appeared that letters were being used to represent a localisable centre or core within a region rather than simply identifying the instrument being represented.

Another departure from the agreed use of the language was in the apparent application of multiple functions to individual descriptors. Several listeners used the FOS descriptor to represent both a ‘feeling of space’ within a vehicle and a region of sound where less definable, less obvious boundaries may have precluded the use of an AREA descriptor. The FOCAL descriptor also appeared to have two related but distinct roles: The first, as expected, representing a localisable centre or core within a less focused region of sound and a second describing a focused, localisable instrument.

Listeners who assigned multiple functions to the graphical descriptors tended to be those who were involved in the development of the language. Since these listeners also demonstrated they could use the inter-subjective descriptors as agreed, multiple roles for the descriptors appeared to result from listeners developing a more complex system for describing their experiences, rather than from their non-comprehension of how to use the existing descriptors. Furthermore, applying multiple roles to the same descriptor may have improved listeners’ representations of their experiences. For example, by describing one instrument within the ensembles using a FOS descriptor towards the rear of the vehicle, and the remaining instruments using the more rigid AREA descriptor at the front of the vehicle, one listener appeared to structure and describe his experiences better in this investigation than in previous studies.

Considering that a researcher tasked with understanding a listener’s experiences can only ever base their comprehension on the listener’s descriptions and not the actual experiences themselves, any individuality in U-GAL use could complicate this task. In the current investigation, when listeners used graphical descriptors as agreed, the task of the researcher was reasonably simple; as the descriptors used by the listener could be understood as representing certain experiences. But, in situations where listeners deviated from using descriptors as agreed, the researcher’s task became less and less transparent until it was no longer possible to comprehend a listener’s experience using a knowledge of how the descriptors should have been used. For example, the increased role of the letter descriptor did not greatly affect the researcher’s comprehension of the listeners’ experiences since letters appeared to have been used similarly to FOCAL descriptors. However in situations where FOS and FOCAL descriptors were used together but without an AREA, it was less likely that the listeners’ individual experiences were being comprehended by the researcher: In essence, were listeners using the FOCAL...
descriptor to represent the region and the FOS to represent a feeling of space around this? Or, conversely, did the descriptors indicate a ‘fuzzy’ region around a localisable centre? Thus when a listener chose not to use the developed descriptors as agreed, more interpretation was required by the researcher. And, without returning a completed descriptive account to the listener for validation, it was impossible to state with confidence that the researcher’s understanding of the listener’s experiences was representative of the listener’s auditory spatial experiences.

Although less ambiguity would be possible if listeners were limited to using the universal descriptors as agreed, this restriction could also present problems for the language’s validity at the start of the descriptive process when asking listeners to represent their experiences. Consequently, this restriction could have a knock-on effect on the provision of a satisfactory answer to the first question posed at the start of this chapter\(^\text{21}\) and the attainment of the research objective. If, for example, listeners were only able to use the AREA descriptor to represent the region covered by a sound, they would no longer be able to describe those more focused sounds perceived as not occupying an AREA. This restriction would also prevent listeners representing instruments experienced as occupying a more fuzzy or cloudy region rather than one possessing a noticeable boundary.

Thus, at this early stage of language development, it was considered unwise to restrict listeners in their use of descriptors. Rather, a process of modifying and further evolving the embryonic language was held to be more appropriate to ensure that any developed language would enable the structuring, representation and communication of a listener’s auditory spatial experiences.

Regardless of the apparent individuality in U-GAL use, it was still possible to evaluate the descriptive language for its suitability as a medium through which individual listeners could describe differences in their experiences. This evaluation is detailed in the following section.

\(^{21}\) whether individual listeners would be able to use the U-GAL in place of their own individual language to structure and represent their individual auditory spatial experiences
4.4 Evaluation of U-GAL using statistical methods

The aim of the statistical analysis was to establish the suitability of the universal-GAL for describing individual listeners’ experiences by testing the hypothesis outlined in the introduction to this chapter, namely:

If a language of universal descriptors has sufficient meaning for a listener - in other words if U-GAL is a suitable medium through which a listener may communicate their auditory spatial experiences - the listener will be able to use the language to represent differences in reproduced audio stimuli where differences in experiences exist.

The first step in conducting the statistical analysis was to identify the elements of a graphical response which could be compared in order to establish the suitability of the language for describing listeners’ experiences: In other words, where should statistically significant differences occur within a listener’s responses?

Physical differences existed between the various items of solo instrument stimuli when presented to the listeners. Both of the percussion and cello recordings used amplitude panning to place the solo instruments at two different locations - centre and left of centre. It was therefore feasible that listeners would vary their placement of these solo sources when represented using graphical descriptors. Additional differences were anticipated between the region covered by the narrow and wide ensembles. The reason for this expectation was that the trio of instruments in the narrow ensemble had been positioned at the centre of the two-channel stereo scene when this stimulus had been created and the same instruments had been amplitude panned to the left of centre, central and right of centre for the wide recording.

At this time it should, once again, be explicitly stated that a statistical analysis of the differences in the stimuli was not considered a sufficient measure of a listener’s ability to use the inter-subjective descriptors. For although differences existed between the various musical stimuli, a lack of difference in listeners’ graphical descriptions could indicate that these physical differences were not experienced, rather that listeners were unable to describe the various stimuli using U-GAL. Thus, when significant differences could not be identified between stimuli, a graphical analysis of visual plots was also conducted to facilitate a greater understanding of the listeners’ responses. This graphical analysis was also an important indicator of a listener’s consistency when repeatedly describing the same stimulus.

4.4.1 Preparing the data for statistical analysis

A consideration for the statistical analysis was how to measure the differences in the position and region covered by the various stimuli. Furthermore it was necessary to identify which U-GAL descriptor (or descriptors) should be used to facilitate this measurement. Providing answers to these considerations was not as straightforward as initially believed. The review of graphical depictions had shown listeners to be using the universal language in different ways. Because of this, an analysis based
on a measure of (for example) the position of AREA descriptors (use to represent the size of a sound) would not prove equally appropriate for all listeners, as not all had chosen to use the AREA descriptor when describing a solo instrument’s position. It was therefore necessary to identify which descriptors would be most effective for highlighting individual listener consistency when using the language.

**Establishing an instrument’s location**

One constant in the listeners’ use of U-GAL was in the placement of letters within either an AREA, FOS (feeling of space) or FOCAL descriptor to represent a stimulus. An examination of listeners’ responses further established that the majority of listeners had positioned these descriptive letters with a degree of forethought, rather than randomly as expected. In situations where a FOCAL descriptor was present, most listeners positioned letter(s) within this descriptor, and when not present, listeners appeared to be positioning letters in place of the descriptor. As the FOCAL descriptor was developed to describe a localisable centre or core within a sound - and since letters were present in all listener responses - differences in the locations of these letters were measured to evaluate individual listener consistency when describing the solo instruments. For all but two listeners this measurement was straightforward and involved determining the left-to-right position of the centre of each letter used to describe a stimulus. However, for listener 7 (where multiple letters had been used to describe the same stimulus) an average position was calculated from all the letters used, and - as responses for listener 8 showed there to be little logic in descriptive letter placement - it was decided to base any statistical analysis on two measurements for this listener; the average position of all letters used to describe a source, and a measure of the central position within the FOS or AREA descriptor used to represent the region covered by the instrument. In the figures and tables which accompany this analysis these two measurements are known as 8a and 8b: Centre of FOS or AREA descriptor and the average letter position respectively.

**Obtaining a measure of the region covered by an ensemble**

Deciding which descriptor to use when assessing the region covered by the two different ensembles proved complicated. Letters representing instruments were not able to change size and were therefore not a suitable means of differentiating between ensembles. Further examination of listener responses indicated that a possible solution would be to measure the region covered by the descriptor used to describe each ensemble, be this an AREA, FOCAL or FOS descriptor. However, a further problem (one specific to ensemble sources) was identified when the graphical analysis of listeners’ responses was conducted: As each instrument within an ensemble could occur within its own distinct region - its own ‘audio image’ or ‘separate sound’ - more than one descriptor could be used to describe the same ensemble. Furthermore these ‘separate sound’ regions could overlap if instruments were perceived to occupy the same location and yet have their own distinct boundaries. Combining measurements of all the descriptors used to represent the region, or regions covered by one ensemble could therefore lead to a numerical measurement up to three times larger than the region actually represented in a listener’s graphical response. An additional ambiguity presented itself when a FOCAL descriptor was used within a FOS: Was the FOS describing the region covered by a ‘separate sound’ or should a
measurement be taken of the FOCAL descriptor and the FOS descriptor disregarded as representing only a ‘feeling of space’ around this more defined region? The degree of interpretation required in obtaining a numerical measure of an ensemble’s region was deemed unsatisfactory for the provision of conclusions which could be said to communicate the listeners’ experiences. To maintain the correlation between listener’s response and researcher’s comprehension, conclusions already obtained from the examination of visual plots were favoured over those obtainable from a statistical analysis of differences. Thus the statistical analysis concentrated on evaluating the U-GAL’s suitability for describing differences in the location of solo instrument sources.

4.4.2 Evaluating differences in the description of instrument location

After measuring the location of each letter used to represent the four solo instrument sources, this data was divided so that responses could be analysed for each listener separately. Dividing the data meant that 12 locations would be analysed for each listener; three per stimulus. Although not by itself a sufficient measure of the language’s suitability for representing a listener’s experiences, the statistical analysis was required to identify where differences between the physically different stimuli had been described by listeners using U-GAL. Differences in description were only considered significant when there was a less than 5% likelihood that listeners had positioned the different sources at different locations by chance alone.

Before a statistical analysis could be conducted to ascertain if significant differences existed between the letter locations, the normality of the data distribution for each item of programme material was established. Results from the normality analysis identified that data from five listeners were found to have a non-normal distribution. Since only three locations existed for each item of programme material, parametric assumptions were violated by this non-normal data distribution. Consequently both non-parametric (Kruskal-Wallis) and parametric analysis of variance (ANOVA) were conducted to identify if differences between solo instrument source locations were significant. Results from the non-parametric analysis concurred with the parametric results and since more information could be obtained about differences between factors using an ANOVA, this latter method is referred to in subsequent paragraphs.

Explaining variance within the ANOVA model

Pertinent results from the analysis of variance (ANOVA) table are included in table 4.D.1 in appendix 4.D. Within the ANOVA, Eta squared ($\eta^2$) was calculated to establish how much variance could be attributed to differences between the four solo instrument sources and how much (the remaining percentage) was due to an external factor. $\eta^2$ represents the squared correlation between the independent variable and any (dependent) variable influenced by the controlled manipulation of this independent variable (Howell, 2002): For example the correlation between a listener’s placement of a descriptive letter (dependent variable) and the manipulation of the programme material (independent

22 Although two dimensional co-ordinates were present for each letter, since only the left-to-right position of the programme material was altered, only the left-to-right position of each letter was used in the analysis.
variable). The closer $\eta^2$ is to '1', the greater the correlation between variables and the more likely it is that variance can be attributed to the independent variable. As the correlation between the variables decreases, variance in the ANOVA model is less likely to be due to the manipulation of the independent variable and more likely to be a consequence of an external factor; for example an increase in listener inconsistency when repeatedly describing the same stimulus. Figure 4.D.1 in appendix 4.D highlighted that, for all but two listeners (listener 16 and 8a) a greater proportion of the variability in the ANOVA model could be attributed to the independent variable ‘stimuli’ than to any other factor. However it was not sufficient to assess the suitability of the language on the values achieved for $\eta^2$ alone, as will be demonstrated in the following paragraphs.

**Identifying significant differences between factors**

In addition to providing information about $\eta^2$, the analysis of variance table (table 4.D.1), provided basic information about listeners’ responses to the different items of stimuli. In particular, the ANOVA identified that of the 22 listeners only three were unable to describe differences between the various external stimuli using U-GAL: These were listener 8: measurement a ($F(3, 8)= 1.997, p= .193$) & b $^{24}$ ($F(3, 8)= 3.224, p= .082$), listener 16 ($F(3, 8)= .397, p= .759$) and listener 19 ($F(3, 8)= 2.780, p= .110$).

For all listeners there existed the possibility that significant differences were being described between some of the stimuli and not others. ‘Post hoc’ comparisons were therefore completed as part of the ANOVA to ascertain where listeners were able to differentiate between the locations of the various stimuli with statistical significance. To perform the most appropriate comparisons, it was first necessary to establish whether variances for all stimuli were equal. Levene’s test confirmed that for ten listeners$^25$, variance was not homogeneous and consequently, both Bonferroni (equal variance assumed) and Games-Howell (equal variance not assumed) comparisons were executed. Results of the post hoc tests, summarised in table 4.D.2 of appendix 4.D, revealed that not all listeners described the solo instrument sources with sufficient consistency to differentiate between the different stimuli. The best results were achieved for the two percussion sources, with 18 of the 22 listeners able to differentiate between the left of centre and central percussion. 13 listeners were unable to describe significant differences between the locations of the two cello sources. The fact that listeners showed a greater capacity for describing differences between the locations of two percussive sources than the cello stimuli was not entirely unexpected as listeners in earlier studies$^{26}$ had mentioned how the cello was difficult to localise. It should be noted however, that there were no obvious differences in size when listeners described the region covered by the solo percussion and cello stimuli (see appendix 4.B,

$^{23}$ Results for listener 8a are those where the centre of a FOS or AREA descriptor was used as a measure of source location for listener 8
$^{24}$ Results for listener 8b are where the average position of letters used in the description of one stimulus was used as a measure of source location for listener 8
$^{25}$ Levene's test on responses for listeners 8b (p = .041), 9 (p = .013), 11 (p = .019), 12 (p = .013), 14 (p = .023), 15 (p = .013), 16 (p = .010), 17 (p = .010), 18 (p = .008) and 22 (p = .014) indicated unequal variance existed between stimuli.
$^{26}$ See Chapter 2, section 2.3.2.
individual listener consistency plots, all pages). The lack of difference between stimuli - when universal descriptors were used to describe the region they covered - reflected descriptions obtained when listeners were asked to describe the same instruments using their own individual graphical languages (see chapter 3, appendix 3.B), with listeners once again describing the percussion and cello sources as occupying the same sized regions within the vehicle.

Since an inability to differentiate between instrument location did not automatically reflect a listener’s inability to use U-GAL descriptors, graphs of means and 95% confidence intervals were prepared to provide a method of visually assessing the results of non-significant comparisons. When compared with the plots of listener consistency from appendix 4.B, these graphs (appendix 4.D, figures 4.D.2-4.D.15) assisted in the explanation of listener responses.

Listener 13 (page 4.B.3 plots 9-12), Listener 1 (page 4.B.2 plots 1-4), Listener 22 (page 4.B.2 plots 9-12), Listener 15 (page 4.B.6, plots 9-12)

Means and 95% confidence intervals for listener 13 (appendix 4.D, figure 4.D.2) identified this listener as one of the thirteen who were unable to differentiate between the two cello sources. When figure 4.D.2 was inspected alongside listener 13’s consistency plots (see figure 4.4.1), a left-bias was revealed in the description of the central cello source. Listener 13 was not alone in experiencing or representing instruments away from their recorded locations. Appendix 4.D figure 4.D.3 (and figure 4.4.2 below) illustrated how listener 1 described sources other than the central cello with a right bias. Furthermore graphs of means and 95% confidence intervals for listeners 22 and 15 (appendix 4.D; figures 4.D.4 and 4.D.5 respectively) coupled with an inspection of visual plots (see figures 4.4.3 and 4.4.4), indicated that both listeners positioned the left-of-centre cello towards the centre of the vehicle. For the above listeners, a shift in the described location of the left-of-centre cello instrument, or the central cello source as exemplified by listener 13, provided a plausible explanation for the lack of significant differences between the left-of-centre cello and the central sources.

Figures 4.4.1 – 4.4.6 Ambiguities in descriptions of the solo cello stimuli

<table>
<thead>
<tr>
<th>Figure 4.4.1</th>
<th>Figure 4.4.2</th>
<th>Figure 4.4.3</th>
<th>Figure 4.4.4</th>
<th>Figure 4.4.5</th>
<th>Figure 4.4.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>L13 (cello C)</td>
<td>L1 (cello L)</td>
<td>L22 (cello L)</td>
<td>L15 (cello L)</td>
<td>L17 (cello C)</td>
<td>L18 (cello C)</td>
</tr>
</tbody>
</table>

27 Listeners who participated in the clarification investigation are known by the same listener number in this evaluation study. Therefore graphical responses in appendix 3.B and 4.B may be compared for the same listener to establish how individual (3.B) and universal (4.B) GAL were being used.
Neither listener 17 nor listener 18 were able to differentiate between the two left of centre sources and the central cello. For both listeners, locating the central cello instrument resulted in the comparatively wide confidence intervals visible in figures 4.D.6 and 4.D.7 of appendix 4.D. These intervals suggested the listeners had described this source with less consistency than the other instruments. An inspection of both listeners’ solo instrument graphical plots ascertained that each listener positioned one description of the central cello source to the left of centre (see figures 4.4.5 and 4.4.6). This one misplacement increased the level of inconsistency associated with the source and prevented significant differences being detected between the central cello and left of centre instruments. One further listener unable to differentiate between cello source location was listener 9. Although from figure 4.D.8 in appendix 4.D it appeared that listener 9 was differentiating between the four solo instrument sources, a paired comparison of cello representations indicated that differences in the locations of these two stimuli were marginally insignificant for this listener (p = .062)\textsuperscript{28}. As with listeners 17 and 18, the misplaced description of one source on one occasion (when describing the left-of-centre cello) provided a likely reason for listener 9’s inability to differentiate between the two instruments.

Even though the above seven listeners did not describe differences between the locations of all of the stimuli, they were able to demonstrate a level of consistency in their representations sufficient to differentiate between some sources. Rather than suggest these listeners were unable to use U-GAL descriptors to represent their experiences, the ability to describe some sources (coupled with rational explanations for why they were unable to describe other sources) suggested that these listeners may have been using the inter-subjective descriptors appropriately, but describing alternative experiences. One possible explanation for these alternative descriptions could have been the stimuli.

The Influence of the stimuli and listening location

In most situations where the above seven listeners did not describe differences between the solo instrument stimuli, it was a cello stimulus which was either moved completely or misplaced when described in one representation. When reproduced within the confines of a vehicle, the low frequency content and continuous nature of the cello sources could have provided ambiguous cues as to this instrument’s location as a result of low frequency standing wave being set up within the vehicle\textsuperscript{29}; something which was less likely to occur for the higher frequency, transient, percussion instruments. However, since both instruments (the percussion and cello) had been amplitude panned to their central and fully left positions, psychoacoustic localisation theory (and in particular the precedence effect\textsuperscript{30}) would suggest that the listening position on the right of the vehicle would have had an influence on both sources; the listeners’ physical proximity to the right of centre loudspeaker meaning that any

\textsuperscript{28} It should be noted that data for both cello left and cello centre were non-normally distributed for listener 9. Although both non-parametric and parametric analyses returned the same results (that significant differences existed between the solo instrument sources) and post hoc comparisons are normally robust to small deviations from normality (Field, 2000, p275), there remains the possibility that a post hoc comparison of these two stimuli was inappropriate.

\textsuperscript{29} See Rumsey and McCormick (1994, p20) for further information about standing waves.

\textsuperscript{30} A useful summary of the precedence effect is included in Rumsey (2001, p28)
signal from this loudspeaker would be perceived earlier than that from the left of centre loudspeaker. Consequently, although stimuli had been amplitude panned to the centre and left of centre, the perceived location of the sources could have been influenced by the earlier arrival time of the signal from the right loudspeaker, potentially drawing the instruments back towards the listening position. The overlaid plots of appendix 4.C demonstrate how the off-centre listening position may have caused listeners to shift descriptions of all instruments towards the right of the vehicle.

Appendix 4.C, page 4.C.6, plots 1-12
Plots 1-12 on page 4.C.6 of appendix 4.C display the different instruments within the four ensembles when described by all listeners using the FOCAL descriptor. When, in the recording, the percussion was positioned at the right of the ensemble, all listeners represented this localisable source in front of their listening location (illustrated in plot 6 on page 4.C.6). In the ensemble where the percussion instrument was amplitude panned to a position mid-way between the centre and the right of the ensemble, there was a pull towards the listening position (see plot 8, page 4.C.4). Finally, when the percussion moved to the centre of the ensemble, greater variation could be seen in listeners’ placement of this source around the centre of the vehicle (illustrated in plots 2 and 12, page 4.C.6). The combined influences of listening location and stimuli on listeners’ responses were just as visible for the voice within the four trio ensembles. Although there was generally more variability when listeners depicted the voice than when they described the percussion, listener consistency was greatest for the voice when amplitude panned to a position in front of the listener (page 4.C.6, plot 9). More variability was introduced when the same source was located at the centre (plot 3 from the same page) or on the left of the ensemble (plots 4 and 10, page 4.C.6) and a noticeable shift to the right of the vehicle could be seen in the listeners’ placement of the left of centre source.

Differentiating between letters positioned outside of FOCAL descriptors
Listener 6 (page 4.B.4, plots 1-4), Listener 22 (page 4.B.2, plot 9)
Listener 6 was a further listener who failed to differentiate between the left of centre instruments and the central cello source. An examination of solo instrument consistency plots (for example figure 4.4.7) highlighted how listener 6 was positioning letters both inside and outside of FOCAL descriptors when describing the different sources. The means and 95% confidence intervals of figure 4.D.9 in appendix 4.D showed listener 6 to be most consistent when describing the central cello source and, for this instrument (figure 4.4.8), letters were positioned within FOCAL descriptors. Thus letters positioned outside of FOCAL descriptors may not have represented anything for this listener. Consequently non-significant differences may have reflected the measure used in the statistical analysis rather than the listener’s use of the U-GAL descriptors themselves. Indeed, when FOCAL descriptors were examined in place of letter descriptors, consistency in the placement of these descriptors was identified. Furthermore differences were visible between the locations of FOCAL descriptors for the various solo instruments. Listener 22 also positioned letters outside of FOCAL descriptors when describing the percussion left source. Post hoc comparisons for this listener, although close to significant at p = 0.052, could not detect any differences between this source and the central percussion instrument. Since this
listener used U-GAL descriptors as anticipated throughout the investigation it is possible that little thought was given to where the listener positioned letters outside of FOCAL descriptors and once again a measurement of letter position was inappropriate.

**Differentiating Between Close Sources**

**Listener 5 (page 4.B.9, plots 1-4)**

Post hoc comparisons and means and 95% confidence intervals for listener 5 (figure 4.D.10 of appendix 4.D) indicated that no significant differences existed between responses when this listener described the location of the central cello source and the left-of-centre percussion and cello instruments. Solo instrument consistency plots for listener 5 (including the left-of-centre cello in figure 4.4.9 above) illustrated that both left-of-centre sources were being represented towards the centre of the vehicle. When the influence of the right-of-centre listening location on both left-of-centre amplitude panned sources was once again considered, a possible explanation could be offered for the non-significant differences in the depicted locations of the various instruments. Furthermore, since all sources were being reproduced to the left of the listener, it was possible that any ambiguity in listener 5’s description of the locations of the instruments reflected the reduced ability of individuals (in general) to accurately judge the location of a stimulus when not positioned directly in front of them.\(^{31}\)

**Listener 19 (page 4.B.10, plots 1-8)**

The results of the ANOVA, coupled with the overlapping confidence intervals of appendix 4.D figure 4.D.13, highlighted that no significant differences were being described by listener 19 when representing the locations of the various solo instrument stimuli. As with listener 5 (figure 4.4.9) listener 19 (figure 4.4.10) positioned all solo sources towards the centre of the vehicle. Although this central positioning could be interpreted as a demonstration of U-GAL’s unsuitability for describing listener 19’s experiences, an inspection of listener 19’s descriptions of the trio ensembles (for example figure 4.4.11) suggested otherwise. When describing the locations of the different ensemble instruments, listener 19 positioned FOCAL descriptors (and individual letters within these descriptors) at different locations within the vehicle; described locations being concordant with the instruments’

\(^{31}\) Although providing a plausible reason for ambiguity in description when listeners remained stationary in their seat, Moore (1997, p227) states that any head movement would reduce localisation confusion.
locations in the physical stimuli. Thus it may be suggested that listener 19, in accordance with many other listeners, experienced (and consequently represented) little difference between the locations of the various solo instrument stimuli.

**Listener 10 (page 4.B.1, plots 9-12)**
Insignificant differences were also reported for listener 10. An experienced listener and previous GAL user (who employed the universal descriptors as agreed), listener 10 once again positioned the letters representing the left of centre programme material and the central cello stimulus very similarly. Figure 4.D.11 in appendix 4.D illustrated how, for this listener, a left of centre bias when describing the central cello could have prevented significant differences from being detected when coupled with the more central positioning of the percussion and cello left. The lack of described differences between the central cello source and the left of centre instruments prevented significant differences from being detected even though listener 10 was more consistent when describing the cello than the central percussion source; highlighting the reciprocal nature between listener consistency and described differences.

**Listener 16 (page 4.B.7, plots 9-12)**
An examination of listener 16’s graphical description alongside figure 4.D.12 in appendix 4.D confirmed that this listener was able to position all but the left-of-centre cello stimulus (figure 4.4.12) with a high level of consistency. Further inspection of the listener’s depictions, coupled with the results of the analysis of variance, established that listener 16 did not differentiate between the locations of the different solo instrument sources, with all instruments located at the centre of the vehicle. For the majority of listeners, non-significant differences when describing the various solo instrument stimuli appeared to be a result of the listening position rather than the unsuitability of the inter-subjective descriptive language. However, for listener 16 this explanation looked less plausible. An examination of ensemble descriptions for this listener (for example figure 4.4.13) indicated that nothing was being represented to the left of centre. Although the description of all stimuli towards the right of the vehicle could suggest a possible influence of the listening position, listener 16 was the only listener not to depict anything to the left of centre for all stimuli. When the lack of difference between listener 16’s representations was coupled with the listener’s alternative use of the universal language (the use of both FOCAL and FOS descriptors independently of AREA descriptors) it appeared more likely that this new listener was unable to describe their experiences using U-GAL, or had not experienced any differences between the various stimuli.
Representing alternative experiences

*Listener 8 (page 4.B.8, plots 9-12)*

When U-GAL descriptors were used to represent the experiences of all except listeners 8, 16 and 19, a level of consistency was demonstrated which enabled the detection of significant positional differences between at least one pair of solo instrument sources. Where listeners (including listener 19) failed to differentiate between sources, rational explanations could be offered and little evidence was presented which suggested listeners were simply unable to use U-GAL descriptors to represent their experiences. But, as with listener 16, listener 8's responses were concerning as no significant differences existed between descriptions of any solo instrument sources (see figures 4.D.14 and 4.D.15 in appendix 4.D) and because this listener chose to use the U-GAL descriptors differently from the remaining participants. Figure 4.D.15 compared mean locations of all letters used by the listener when representing the different sources. This graph revealed how listener 8 was more consistent when positioning the left-of-centre instruments (for example the left-of-centre cello in figure 4.4.14) than the central sources (figure 4.4.15). Only one letter was used by listener 8 when describing each repeated presentation of the left-of-centre sources, but multiple letters were used to represent the central sources.

By using multiple letters, it was possible that listener 8 was not representing anything specific and consequently a statistical analysis based on letter locations was flawed for the central instruments.

Further light was shed on listener 8's inability to differentiate between sources by their answer to one of the post-investigation questions\(^{32}\). When asked for any additional comment, the listener stated that it was "very easy to describe a sound's position and spaciousness". The use of the term spaciousness suggested that listener 8 may have been representing an alternative experience using U-GAL descriptors and indicated why the listener favoured the FOS descriptor when representing the region covered by each solo instrument and had not used any FOCAL descriptors. Furthermore, the listener's comment, coupled with the physical properties of the different stimuli, provided an explanation as to why no differences existed in the left-to-right extent of each solo instrument description. Since a small amount of reverberation had been added equally to all stimuli, if listener 8 had in fact described the spaciousness of the reproduction using the FOS descriptor rather than the location and size of the various stimuli no differences would be expected in the resultant representations.

\(^{32}\) See appendix 4.E for a full list of questions asked in the post-investigation questionnaire and listeners' responses
4.4.3 Evaluating the use of U-GAL by different listening groups

Listeners 8 and 16 were included in the evaluation study as examples of listeners with no previous experience of using graphical descriptive methods or developing the U-GAL. Considering naive listeners were included to provide a measure of the external validity of the language, both listeners (seemingly alternative) use of the graphical language (with listener 8 moulding descriptors to fit their own experiences and listener 16 representing few differences with the language) suggested that naïve listeners may have been unable to use the developed descriptors to describe their experiences and indicated a possible limitation of the language. As a result of this, a specific analysis was conducted to establish whether the eleven naïve listeners were able to use the language developed by the remaining listeners in the study. This specific evaluation involved comparing responses obtained from both new and existing listeners, testing the theory that: if U-GAL provided all listeners with descriptors suitable for representing their experiences (and all listeners could understand their task within the investigation) no obvious differences would exist between naïve and existing listeners’ responses.

Descriptions of the four solo instrument sources were once again compared using the location of the letter descriptor as a measure. Figure 4.D.16 in appendix 4.D, a graph of means and 95% confidence intervals for this comparison, highlighted that both listening groups responded to the different items of stimuli with a similar level of consistency. Furthermore since the confidence intervals for listening groups overlapped, it was possible to state that both groups were describing the locations of the different sources with no obvious differences. Consequently, when naïve listeners in the study described the locations of the different stimuli, it was determined that U-GAL possessed a degree of external validity.

As revealed in section 4.2.2, regardless of whether or not they had been involved in the development of the U-GAL, a participant in the evaluation could be a trained listener, an experienced musician, or inexperienced in both respects. Consequently a further analysis of listeners’ responses was conducted to establish if prior experience - previously defined as any training in listening either as a musician or more specifically as a listener trained in the evaluation of reproduced audio - influenced language use. Descriptions of solo instrument location were, once again, compared but this time for two different levels of listening experience; some experience or no experience. From the means and 95% confidence intervals of figure 4.D.17 in appendix 4.D, a listener’s previous experience of playing a musical instrument, or their training in listening, was seen to have little bearing on their use of U-GAL’s letter descriptor. Furthermore, a listener’s consistency when describing the different sources was not influenced by this listening experience. Results suggested that a listener could use the inter-subjective descriptors to represent the location of a stimulus regardless of whether or not they had received any musical or listening training.
4.4.4 A brief comparison of listener descriptions

Because U-GAL was developed to represent an individual listener's experiences, each listener's use of the language was evaluated individually. Other than a brief analysis of how different listening groups were using the language, differences in the use of descriptors by different listeners were not formally assessed using statistical methods. That said, figures 4.D.18 - 4.D.21 in appendix 4.D provided an indication of how all listeners were describing the same stimulus. In these graphs, means and 95% confidence intervals of the left-to-right positions of the solo sources were plotted independently. By comparing means and confidence intervals within the same graph it was possible to identify similarities in listeners' positioning of the same stimulus and to compare listeners' consistency. Thus an individual listener's responses could be put into context. For example, despite representing sources other than the central cello with a right bias, the consistency of listener 1 when describing the cello left was better than that of other listeners when representing the same instrument and suggested that, even if different, listener 1 was confident in their placement of this instrument at the centre of the vehicle.

Thus, even though between-listener analysis could not provide a fail-safe measure of U-GAL's suitability (as listeners could be representing alternative experiences to one another) this comparison was still informative. By way of further illustration, an examination of figures 4.D.18 - 4.D.21 in appendix 4.D, alongside figures 4.4.16 - 4.4.19 below (and FOCAL descriptor plots on page 4.C.2 of appendix 4.C) identified that consistency between listeners was greater when describing the two central sources (for example figures 4.4.17 and 4.4.19 and appendix 4.D figures 4.D.19, 4.D.21) rather than the two left-of-centre sources. As listeners appeared consistent when placing the central instruments at the centre of the vehicle, it was less likely that the lack of consistency (between listeners) when placing the left-of-centre sources resulted from their inability to use U-GAL than their representation of different experiences. Once again, an explanation for the between-listener differences could have been the right-of-centre listening position pulling the left-of-centre images towards the listener.

Figures 4.4.16 - 4.4.19 Letter locations plotted for all listeners

Figure 4.4.16 (perc L)  Figure 4.4.17 (perc C)  Figure 4.4.18 (cello L)  Figure 4.4.19 (cello C)

33 source location was once again established by measuring the position (left-to-right) of the letter descriptors used to represent each solo instrument.
4.5 Evaluating a listener’s ability to use U-GAL – summary of findings

The aim of the combined statistical analysis and graphical analysis of visual plots was to establish the suitability of the universal-GAL for describing individual listeners’ experiences; in essence identifying whether listeners were able to use the inter-subjective graphical descriptors to represent individually perceived differences in various audio stimuli. The hypothesis tested was:

If a language of universal descriptors has sufficient meaning for a listener - in other words if U-GAL is a suitable medium through which a listener may communicate their auditory spatial experiences - the listener will be able to use the language to represent differences in reproduced audio stimuli where differences in experiences exist.

The combined evaluation of visual plots from appendix 4.B and 4.C and results of the statistical analysis of variance (appendix 4.D) identified that, regardless of listening experience or any prior knowledge of using or developing a graphical language, 19 of the 22 listeners were consistently able to represent significant differences between at least two solo instrument stimuli when describing source location using U-GAL descriptors. Within this subset of listeners, eight were able to differentiate between the locations of all the source items. For these eight listeners it could be concluded that a language of universal descriptors provided a suitable, reliable medium via which they could describe the perceived location of different stimuli.

Although the remaining listeners did not differentiate between all solo instrument stimuli, it was possible that the lack of differences in their representation reflected a lack of difference in their experiences, or some unwanted factor introduced by the investigation, rather than an inability to use the graphical language to describe the stimuli. Thus, when listeners failed to describe significant differences between sources, the evaluation process attempted to explain these occurrences using individual listener consistency plots from appendix 4.B and the means and 95% confidence interval graphs of appendix 4.D. Using these additional methods, several potential reasons for the description of non-significant differences were identified. These explanations can be summarised as follows:

4.5.1 Explanation 1: The influence of programme material and listening location

The most common response which led to non-differentiation between stimuli occurred when listeners positioned the left-of-centre percussion or (more usually) the left-of-centre cello towards the vehicle’s centre. Also associated with this action was a listener temporarily misplacing a source; describing the same instrument in a different location for one repeat. Any perceived shift in location, towards the centre of the vehicle by a left-of-centre source, meant that greater descriptive consistency was required in a listener’s responses in order for significant differences between stimuli to be revealed. Similarly, one misplaced descriptor caused the 95% confidence intervals for the stimulus to widen, and potentially prevented significant differences from being identified between that stimulus and the remaining instruments.
Where listeners misplaced stimuli, or positioned left-of-centre and central stimuli together, it was conceivable that the listening position on the right of the vehicle was affecting the listeners' auditory perception of the spatial scene, influencing their description of the solo instrument stimuli. Since these stimuli were amplitude panned to the centre and left of the vehicle, it was likely that the listeners' proximity to the speaker on the right of the vehicle caused the amplitude panned stimuli to shift to the right or to be ambiguously located when experienced. Thus, the effect appears to be one of the stimulus and listening location and not of the listeners' inability to describe differences using U-GAL.

In descriptive analysis (DA), participants involved in the development of a precise lexicon for the reliable description and differentiation of food items, firstly discriminate between disparate samples on the developed attribute scales before moving on to describe and discern smaller differences. According to Meilgaard et al. the use of disparate samples at the early stage of descriptor development “allow the panel to see that the terms and scales are effective as descriptors and discriminators” (Meilgaard et al. 1999, p144). Although U-GAL differs from DA in its objectives (with U-GAL allowing for the non-differentiation of different reproduced audio stimuli if this is indeed a listener’s experience) it is possible that a lack of spatial differences between stimuli caused the descriptive ambiguity in listeners' responses. Consequently, one consideration for further work would be that listeners describe their experiences of a more disparate range of stimuli using the inter-subjective graphical descriptors. This use of different stimuli would not only establish a better sense of the ‘difference threshold’ which may be determined using the graphical language, but also provide an indication of the language's external validity; stimuli having remained a constant feature in the development of U-GAL to date.

4.5.2 Explanation 2: The inappropriateness of the statistical measure

Although the letter descriptor was the only descriptor to be used by all listeners when describing an instrument, the position of this letter appeared more meaningful for some listeners than for others. For two listeners who used letters outside of the FOCAL descriptor, and for a further listener who used multiple letters to describe the same source, few significant differences existed when the left-right positions of these letters were analysed. Since the agreed role of the letter descriptor was as a label for the instrument being described, the use of an inappropriate statistical measure appeared a plausible explanation for the non-significant differences in listeners' descriptions where letters had been used as labels rather than indicators of a source’s ‘centre’ or focus.

It is believed that the further evolution of the descriptive language should improve the researcher’s ability to find an accurate statistical measure of a listener’s responses. This belief is held because, in developing the inter-subjective descriptors to better structure and represent listeners’ experiences, there should be more agreement between listeners on how to apply descriptors to experiences. Consequently

---

34 the difference threshold was defined by Meilgaard et al. (1999, p124) as “the extent of change in the stimulus necessary to produce a noticeable difference”.

N Ford  Doctoral Thesis  181
the researcher will be able to base a statistical analysis on a more appropriate, more communal, measure.

4.5.3 Explanation 3: The representation of alternative spatial attributes

The two (naïve) listeners who were unable to successfully differentiate between the locations of any stimuli using U-GAL had chosen to use the inter-subjective descriptors in roles contrary to those agreed when the language had been developed. For one of these listeners it was not possible to ascribe the similarities in their descriptions to anything other than a lack of difference in their experience or an inability to use the graphical language to represent any differences which did exist. For the other listener, the term ‘spaciousness’ was mentioned when providing responses to the post-investigation questionnaire and the feeling of space (FOS) descriptor was used without a ‘centre of sound’ (FOCAL) descriptor to represent the majority of stimuli.

The use of this alternative ‘spaciousness’ term by the listener when describing their responses provides an explanation for the lack of significant differences in the described locations of the different stimuli. If the listener was indeed describing an alternative spatial attribute named spaciousness, the lack of difference in stimuli location may therefore have reflected a lack of difference in the spaciousness of the solo instrument stimuli as experienced by this listener. Since no manipulation of spaciousness had been attempted when the stimuli had been created, a lack of difference in the description of this alternative attribute also reflected the properties of the external stimuli. However, a question remained following the above finding, namely; why did the listener choose to represent spaciousness and not the descriptors mentioned in the instructions for listeners sheet?

Although many of the naïve U-GAL users in the investigation were able to graphically describe experiences which appeared to satisfy the written instruction (suggesting an improvement to previously worded instruction), at this early stage of the language’s development it may have been that listeners were still misunderstanding verbal terminology and mistakenly describing different experiences which could be structured using the verbal terminology used in the written instruction. Alternatively, since previous GAL studies had indicated that a listener would mould unsuitable written instruction to fit their experiences (when unable to relate this to their experiences), the description of spaciousness may have reflected the listener’s actual experience of the stimuli better than a description of area or location.

Thus, the adverse influence of an inappropriate verbal language could not, as yet, be ruled out for listeners who had used the developed descriptors in an ambiguous manner. This finding further highlighted the necessity for the verbal language to reflect the experiences of the listener when structured and represented graphically in U-GAL, and indicated why the continued development of the verbal language should be considered alongside the evolution of the graphical language.

---

35 See appendix 4.E
4.6 U-GAL evaluation conclusions

As a result of the statistical and graphical analyses of listeners' descriptions, the universal-GAL was deemed an unsuitable medium for the description of the left-right location of solo instrument sources for two listeners due to the ambiguous nature of their descriptive responses. The rationale behind this decision was that it was not possible to attribute the lack of differences in these listeners' responses to their individual experiences. Thus the possibility existed that these listeners were not able to structure, represent or communicate their auditory spatial experiences using the descriptive graphical language. However, for the remaining 20 listeners, the combined statistical analysis and evaluation of listener consistency plots suggested that an inter-subjective language (developed through discussion and the communalisation of listeners' experiences) was a suitable medium for describing the spatial characteristics of solo instrument stimuli as experienced by an individual listener.

Although inter-subjective descriptors appeared suitable for describing the majority of listeners' experiences, (thereby successfully progressing from the listeners' experiences to their graphical representation in the descriptive process model), in concluding this chapter it is worth recollecting why the universal graphical communication medium was being developed.

In developing U-GAL it was believed that the language would better enable the researcher to comprehend each listener's experiences. This belief was held as existing listeners within the study had developed each descriptor within the universal language to structure their experiences and to represent specific spatial attributes as perceived. Less interpretation was therefore assumed to be involved for the researcher, since their understanding of a listener's experiences would be based on a knowledge of how and why each descriptor had been developed and the experience it was representing. Essentially, rather than the researcher attempting to understand and communicate a listener's experience described using a listener's own language, both listener and researcher would have the facility of a mutual descriptive language in which to communicate. In developing an accompanying set of verbal descriptors, it was further anticipated that the researcher would be able to communicate in the same language as the listener; one representative of the listener's auditory spatial experiences and the graphical descriptors created to represent these experiences.

Consequently it was assumed that the development of U-GAL and the accompanying verbal descriptor set would enable the researcher to arrive at a valid comprehension of a listener's experiences and, in this way, it was envisaged that steps would be taken towards completing the descriptive process from representation to comprehension and correspondingly, fulfilling the research objective.

37 A formal statistical analysis was only conducted for listeners' graphical representations of solo instrument location, however the graphical analysis considered entire depictions.

38 See figure 1.4.1 in chapter 1 for one instance of the descriptive process model or appendix 1 for a key to the descriptive process
When the universal language was evaluated with respect to this second part in the descriptive process, a graphical analysis of listeners' responses established that even when the universal language was appropriate for individual listener use (U-GAL enabling listeners to describe their auditory spatial experiences) ambiguity could still exist in how the language was being used.

For some of the listeners (capable of using U-GAL to differentiate between physical stimuli) the graphical descriptors were not always being used as agreed, since listeners apparently altered the existing descriptors to represent additional spatial attributes. One common departure from the agreed use of the language was for listeners to use both FOCAL and FOS descriptors in addition to the AREA descriptor to represent the region covered by a stimulus. A possible reason for this use of multiple descriptors was that each 'separate sound' region was not only being described in terms of its size. Instead each sound was being described with respect to its focus; stimuli ranging from very focused (described using the FOCAL descriptor), to fuzzy when the FOS was employed.

This development of the existing graphical language was not restricted to new listeners (a phenomenon which could simply have been attributed to these naive listeners misunderstanding how to structure their experiences using the various universal descriptors) but also involved existing, experienced listeners who were present when U-GAL was being created and were therefore responsible for the agreed use of each descriptor. That even existing listeners appeared to be developing the universal language further should not have been entirely unexpected due to the infancy of the universal graphical medium.

The comparatively early stage of the universal language's development could have influenced listeners' responses in different ways. Prior to the development of U-GAL, listeners had been using their individual descriptive terminology to graphically structure and represent their own experiences. With the introduction of U-GAL, initial steps away from this immediate, individual terminology were being made. However, the likelihood remained that listeners were still coming to terms with the augmentation of their familiar lexicon with the common language. Accordingly, even though U-GAL was developed through the mutual agreement of listeners, these listeners would still have had a comparative lack of familiarity with the common terminology when this was related to their own descriptors. In an attempt to return to more familiar descriptors, listeners could therefore have individualised the inter-subjective descriptors to such a degree that they were no longer recognisable in their universal agreed application.

A competing explanation for the development of the graphical descriptors could be a listener's increasing awareness of the experiences which could be structured and represented using the language. This justification would also explain why the more experienced listeners tended towards developing the language rather than their more naive counterparts: The familiarity of these listeners with structuring their experiences in a graphical language necessitating the development of a more comprehensive descriptor set.
Regardless of the reason, any individual development of the inter-subjective language could cause problems for the researcher attempting to comprehend the listener's experiences through only their knowledge of how and why the inter-subjective descriptors had been developed. A useful example of this phenomenon occurs in the everyday use of language. Each individual has their own language taken from the language system that is available to them. Many different words exist which can be used to describe individual experiences and a greater chance of successful communication occurs when individuals use words from the language system in their defined role rather than alternatively to express their experiences. However, there is a degree of elasticity in how terminology may be used, allowing individuals the flexibility of applying language in their own way to their own experiences. It is only after a certain point that this elasticity causes the communication process to fail; when the experience can no longer be understood.

Accordingly, in the evaluation of U-GAL, the use of descriptive letters rather than the FOCAL descriptor (to represent a focal point within the stimuli) was not overly concerning, since the researcher could still comprehend the listeners' experiences when descriptors were used in this way. However, the potential for ambiguity in the researcher's comprehension (and any subsequent communication) of listeners' experiences increased when listeners used the universal descriptors in a more individual way. This was demonstrated when listeners used both FOCAL and FOS descriptors without an AREA.

When attempting to understand the listeners' experiences represented in this way, the researcher could not be certain which (if either) of the FOCAL or FOS descriptor represented the region covered by a stimulus and consequently ambiguity was introduced. Hence, the researcher's ability to provide a comprehensive account of a listener's experiences decreased as the individuality in language increased. Similarly, the potential for fulfilling the research objective diminished when the language's actual use by listeners was contrary to its agreed use. Less ambiguity would be involved for the researcher if listeners were limited to using descriptors as agreed. However, restricting a language at too early a stage in its development was considered unwise since it may prevent the creation of a U-GAL more representative of listeners' auditory spatial experiences, or restrict listeners to describing relatively naive experiences rather than those which could arise from the further communalisation of experiences.

Instead of restricting listeners to using the existing descriptors as agreed, a process of modifying and evolving the embryonic language was deemed to be more appropriate at this current stage of the research. The process of advancing the descriptive language's development would involve asking listeners to clarify their graphical descriptors in much the same way as previously defined in chapter 3. This clarification process was also deemed advantageous, since the only evaluation of the listeners' ability to use the U-GAL (when significant statistical differences had not been identified between stimuli) had been provided by the researcher. Thus the process of returning responses to listeners would aid in the continued development of the universal descriptive language and its concurrent validation.
Fuller details of the work proposed to further the development of the universal language will be provided in chapter 5. But to better illustrate the current stage of the descriptive language’s development, figure 4.6.1 is included as a model of U-GAL following the evaluation investigation.

In conclusion, for the universal graphical communication medium U-GAL to be considered suitable for the task of communicating an individual listener’s experience (Exp) of a particular auditory stimulus (St), the language should firstly enable the listener to structure and represent (R) their experiences (whatever these may be) and secondly, with the existence of a graphical representation, enable the researcher to comprehend (C) each listener’s experience through this graphical description. Throughout each stage of communication, from experience-to-representation and from representation-to-researcher’s comprehension, the language should communicate the listener’s experience without introducing any ambiguity (indicated by a dotted line in figure 4.6.1) which would prevent the listener’s experiences from being comprehended.
Until the development of the universal graphical communication medium, there invariably existed the possibility for ambiguity in the researcher's comprehension of the listeners' auditory spatial experiences since these individual experiences were being communicated using individual languages. The evaluation of U-GAL demonstrated how, by developing the universal language, there existed the possibility for the researcher to obtain from the listener a comprehensible graphical description without ambiguity when the language was being used to structure and represent listeners' experiences as agreed. This possibility - illustrated in figure 4.6.1 by following a route from the listener's experience (Exp1) through their description of this experience using the universal language (UR) to the researcher's comprehension of the listener's experiences in (CR) - meant that the research objective could be fulfilled when U-GAL descriptors were employed, and that (in these circumstances) the language could be deemed suitable for the task of communicating a listener's experiences. However, as indicated by the multitude of possible routes in the same figure, this desired outcome was not attained for all listeners.

Alternative understandings (CM) of a listener's experience occurred when listeners chose to use the developed descriptors in a manner other than that agreed to describe their experiences (CM1), or when listeners were unable to use the universal descriptors to represent their experiences (CM3 or CM5). When listeners were unable to use the universal descriptors, or were using these inter-subjective descriptors ambiguously, one possible explanation was that listeners were describing alternative attributes to those requested in the written investigation instruction (Q). Thus, there still existed the possibility that the verbal language developed alongside U-GAL was not providing appropriate verbal instruction and consequently listeners were still having to choose between representing their auditory spatial experiences or responding to the written instruction and not describing their experience. It should, however, be noted that no specific evidence of this last eventuality was identified during the evaluation investigation.

It is encouraging that, at this stage in U-GAL's development, an unambiguous progression is possible through the descriptive process from the listeners' auditory spatial experiences to the researcher's comprehension of these experiences when graphically represented. In following this one route, it is possible for the research objective to be fulfilled. Nevertheless, anomalies in U-GAL use still suggest the language to be in its infancy, with further inquiry and development required before U-GAL can be regarded as an effective communication medium.
4.7 Chapter summary

Chapter 4 has been concerned with the process of evaluating the universal graphical language (U-GAL). As a result of U-GAL's development, it was believed that the second part in the descriptive process (the researcher's comprehension of listeners' experiences when graphically represented) would be less ambiguous and, consequently, it would be possible to fulfil the research objective.

In section 4.3 a graphical analysis highlighted how, although many listeners (both new and existing) were using the graphical descriptors as agreed, other listeners appeared to adopt a more individual approach to applying the universal descriptors to their individual experiences. The more listeners deviated from the agreed use of the graphical descriptors, the more the task of the researcher was complicated. Section 4.3 concluded by stating that, although the researcher's ability to comprehend a listener's experiences would be improved by ensuring listeners used graphical descriptors as agreed, such a restriction at this early stage of the language's development could pose a problem for the language's validity, since the restricted language may no longer reflect the listeners' experiences. Consequently it was suggested that the existing descriptors should be further developed to ensure their adequacy for communicating listeners' auditory spatial experiences.

Chapter 4 was also concerned with establishing whether the use of universal (rather than individual) descriptors hindered the listener when structuring or representing the spatial attributes of reproduced audio stimuli as experienced. In section 4.4, it was established that 86% of the listening population were able to differentiate between the locations of at least two items of stimuli, confirming the suitability of the U-GAL. Since a lack of statistical significance could reflect a lack of difference in listeners' experiences as well as an inability to use the graphical language, a graphical analysis was also conducted. Results from this analysis provided explanations where listeners were not able to differentiate between solo instrument stimuli.

The graphical analysis suggested that the choice of stimuli combined with an off-centre listening location may have resulted in listeners failing to differentiate between stimuli. It was also possible that the measure used in the statistical analysis (the location of the letter descriptor representing each solo instrument stimulus) was inappropriate when the placement of this descriptor was random. A third explanation for the lack of described differences was that listeners were representing alternative experiences to those requested; experiences which were not manipulated in the physical stimuli. The description of alternative experiences appeared to be a consequence of listeners either misinterpreting verbal instruction, or moulding the graphical language to fit their actual experiences.

Chapter 4 concluded by stating that the universal-GAL was still in its infancy. And although the language could be used as agreed by many listeners to successfully describe differences in their experiences (thus fulfilling the research objective) there still existed the possibility that listeners would not be able to use the language, or would manipulate the language until they were able to use the
universal descriptors to structure and represent their own experiences. Thus, the potential for miscomprehension by the researcher still existed.

Further inquiry and exploration were proposed to increase the potential for U-GAL to fulfil the research objective. Proposed work included returning the current languages (both verbal and graphical) to listeners for clarification. An overview of this further work is included in chapter 5.
CHAPTER 5

Conclusions

Further work
Chapter 5 - Conclusions

5.0 Chapter overview

Chapter 5 presents the conclusions of the research and outlines a proposal for the further development of the descriptive graphical language U-GAL.

Concluding remarks in section 5.1 summarise the language's development, provide details of the context in which the attainment of the research objective should be considered, and propose why the current method for developing a listener-focused descriptive language can be considered an original contribution of the author to subjective audio evaluation.

Section 5.2 defines how ambiguities present at this stage in U-GAL's development could be minimised by the future evolution of the descriptive medium using the current language-development method. The section concludes by proposing an amendment to the existing development method which allows for the validation of a descriptive language following its evaluation.
5.1 Conclusions

The research presented in this thesis has endeavoured to engage with (rather than screen-out) the subjectivity of the listener, with listeners' individual auditory experiences regarded, \textit{a priori}, as valid. The acknowledgement that each listener has their own personal experience of a reproduced audio stimulus shifts the focus of the subjective evaluation and, rather than the listener attempting to understand what qualities of the stimulus the researcher would like them to measure, the onus is on the researcher to explore and understand the listeners' individual experiences. This exploration ensures that subsequent studies consider attributes relevant to the listener, thereby improving the validity of subjective audio evaluation.

Figure 5.1.1 U-GAL: A descriptive graphical language

To enable the exploration of listeners' individual auditory spatial experiences - and facilitate effective (unambiguous) communication between listener and researcher - a descriptive graphical language (U-GAL, see figure 5.1.1) was developed. As the purpose of the language was to communicate listeners' individual experiences, it was not sufficient for the language's effectiveness to be evaluated by correlating the listeners' use of the developing graphical descriptors with the researcher's experience of the stimulus. Accordingly, in situations where statistical and graphical analyses identified apparent inconsistencies in listeners' responses (or where depicted differences did not correspond with physical differences in the reproduced stimuli) instead of dismissing listeners' responses as outlying, it was necessary for these seemingly anomalous data to be investigated further. The process of going beyond conventional quantitative analysis methods - to focus on systematically identifying and minimising ambiguities in listeners' responses - led to the employment of a novel method for developing the descriptive language. This method can be briefly outlined as follows:

\textbf{Stage 1: Elicitation} of individual descriptors (from naive and experienced listeners) \textit{Chapter 2}

\textbf{Stage 2: Clarification} of listeners' descriptors \textit{Chapter 3}

\textbf{Stage 3: Development} of communal language based on listeners' individual descriptors \textit{Chapter 3}

\textbf{Stage 4: Evaluation} of communal language (using new and existing listeners) \textit{Chapter 4}

At each stage in the language's development, the simultaneous evolution of a descriptive process model provided a means of visually illustrating different (either ambiguous or effective) descriptive routes between listener and researcher. For example, in this thesis, the \textit{clarification} of listeners' individual graphical depictions identified that the terminology used by the researcher when providing listeners with instruction was a source of ambiguity. The descriptive process model therefore included the question (Q) posed by the researcher as a potential source of ambiguity. As a means of minimising this ambiguity, a verbal language was developed alongside the graphical descriptors during stage 3.
The graphical representation of listeners' auditory spatial experience

Chapter 5 - Conclusions

The chronology of developing U-GAL is visualised in figure 5.1.2, with each stage in the language's development represented by one of the rectangular boxes in this figure. The proximity of each box to the bottom right of the diagram (the effective comprehension of listeners' experiences) provides an indication of the language's effectiveness at each stage in its development. For example, although listeners' experiences could be described using the individual graphical languages (I-GALs) elicited during stage 1 of the language's development, the use of these individual descriptors was less likely than the development of U-GAL (stage 3) to result in the ultimate objective of the researcher comprehending the listeners' experiences.

Figure 5.1.2 The language development process

The descriptive language's evaluation stage confirmed that the researcher was able to comprehend listeners' individual auditory spatial experiences when graphically represented using the medium. Consequently, a positive statement could be made regarding the attainment of the research objective. It is, however, prudent to acknowledge where specific aspects of the language's development may have restricted or influenced the direction and conclusions of the work.

5.1.1 Influences of the research context and other considerations

The language presented here has undoubtedly been influenced by the specific context of its development. For instance, different stimuli could give rise to different auditory experiences which would require representation using alternative graphical descriptors. The setting of investigations (within an automotive environment) should also be recognised alongside any declaration of the effectiveness of the graphical language for fulfilling the research objective. Furthermore, although the listening population was extended over the project's duration from three to 31 people and included both experienced and inexperienced listeners (to improve the external validity of the descriptive language), an alternative population might have developed a different set of graphical descriptors.
Although it is feasible to suggest that the same U-GAL could be effective when used in alternative contexts, it would be unwise to generalise the conclusions of this research without further investigation. Currently, the descriptive graphical language can be used for representing the selected listeners' auditory spatial experiences of simple musical stimuli when reproduced over a multichannel audio system within an automotive environment.

Furthermore, although listeners developed their own graphical language to represent their experiences, it is acknowledged that the current language has inevitably been framed from the perspective and understanding of the researcher; it was the researcher who defined the research objective, analysed data and selected the listeners, programme material and the investigation procedures.

Considerations regarding the research ethos
In this thesis, a listener's experiences are regarded as valid. However, in certain circumstances the participation of a particular individual in a subjective study may be inappropriate. Where, for example, the objective is the description of an external stimulus, a listener who does not experience any difference when that stimulus is manipulated (possibly due to a known hearing pathology) may be considered an unsuitable participant. Care should be taken to ensure that a prospective listener is not excluded as a result of anomalies elsewhere in the descriptive process.

Considerations regarding the use of language
With language, it is not possible to obtain an exact replica of an individual's experiences. Rather, what is obtained is a representation of experience in a communicable form. Since language and auditory experience are not expressly linked, it is likely that the medium will have modified the message; with the listeners' experiences moulded to fit the form of their communication. Due to the difficulties of structuring experiences in language, listeners will most likely have concentrated on describing the auditory spatial experiences most readily represented. Accordingly, the use of a graphical medium will have enabled a perspective to be gained on listeners' spatial experiences, but the limitations of this graphically framed perspective are duly acknowledged.

The balance between retaining meaning & obtaining understanding
Quite ironically, in an attempt to understand more about a listener's individual auditory spatial experiences, it has been necessary to employ a universal language, ensuring the accessibility (in communication) of each individual's personal constructs. However, the inter-subjective discussion leading to the development of this universal language will have transformed (to some degree) the listeners' ways of thinking about their experiences, affecting the information graphically communicated to the researcher. Whilst the process of developing U-GAL may have altered the listeners' understanding of their original experiences, these experiences will remain individual to the listener. Consequently, as with any common language (for example the English language), U-GAL should still enable the communication of listeners' individual (if not original) experiences.
5.1.2 Concluding remarks

In conclusion, the systematic development of a descriptive graphical language has enabled the listener to graphically describe their individual auditory spatial experiences in a manner which can be understood by the researcher. Accordingly, when the research conclusions are considered with respect to the original contributions summarised in section 1.6 of this thesis, it is possible for the author to claim that the particular descriptive graphical language developed during the research project provides a context-dependent contribution to subjective audio evaluation by (i) positioning meaning with the listener, (ii) using elicited or developed graphical terminology and (iii) enabling the investigation of perceived spatial characteristics in reproduced audio.

Whereas the validity of a particular language, for example U-GAL, is limited by the specific context of that language’s development, the method for developing a listener-focused descriptive language (by methodically investigating and minimising ambiguities in listeners’ responses) is not limited in this way. This independence allows for the method’s use in other contexts where the development of a listener-focused descriptive language is required. Furthermore, the four-stage elicitation, clarification, development and evaluation process is iterative, ensuring that a developing language can continue to evolve in response to findings from subjective evaluations or changes in context.

Thus, with the completion of the research, the evolution of a novel method for developing descriptive languages is claimed, by the author, as an important contribution to the field of subjective audio evaluation.

Although the current U-GAL enables the effective (if context dependent) communication of listeners’ individual auditory spatial experiences, ambiguities in the descriptive process are still possible (for example, where listeners have used descriptors in a manner contrary to that agreed) suggesting that a further period of language development might be necessary.
5.2 Further development of the descriptive language

It is proposed that improvements to the current descriptive graphical language (U-GAL) are made using the iterative language-development method evolved by the author and described in this thesis. Accordingly, further research is envisaged which comprises the following elements:

(i) **Re-evaluate** individual listeners’ use of the current U-GAL to obtain contemporary information about how listeners are describing their auditory spatial experiences. (It is proposed that re-evaluation employ the method detailed in chapter 4 of this thesis).

(ii) **Clarify** any ambiguities in the listeners’ use of U-GAL by returning graphical depictions to the individual respondents for comment. (The clarification method is described in the first section of chapter 3).

(iii) **Develop** the current descriptive graphical language based on the findings of the clarification stage. Improvements will be made during a period of inter-subjective listener discussion (the method for which is described in the second section of chapter 3).

(iv) **Evaluate** individual listeners’ use of the novel descriptors in a formal investigation.

Since language is context dependent, it is prudent that the effectiveness of U-GAL be re-evaluated when the context in which it is to be used changes. The iterative nature of the language-development method allows for the evolution of U-GAL in this situation or when the evaluation of novel audio devices necessitates the development of the language. Currently, a change in context is proposed with respect to the stimuli employed. Specifically, the listeners’ use of U-GAL to describe a disparate range of stimuli requires investigation. This modification is necessary in order to identify whether the lack of descriptive differences observed during the previous evaluation stage (see chapter 4) can be attributed to insufficient diversity between the programme material.

Even though the iterative nature of the method ensures that a descriptive language can be modified according to context, this doesn’t necessarily guarantee the validity of the researcher’s comprehension. To verify that the researcher understands the auditory experiences being described using the developed descriptors, listeners should be provided with an opportunity to comment on any analysis undertaken by the researcher. Consequently, it is proposed that a fifth stage be incorporated in the language-development method. The final method can therefore be summarised as follows:

**Stage 1:** Elicitation of descriptors from individual listeners

**Stage 2:** Clarification of individual descriptors by listeners

**Stage 3:** Development of common language via a process of inter-subjective discussion

**Stage 4:** Evaluation of common language (including analysis by researcher)

**Stage 5:** Validation of analysis by listeners (to ensure the researcher’s comprehension of listeners’ auditory experiences is adequate)

---

1 Re-evaluation is required due to the time elapsed since the original evaluation of U-GAL.

2 In due course, consideration may also be given to the further development of stimuli with respect to complexity. This proposal is tentative as the use of more complex material could potentially complicate the task of both listener and researcher and introduce a greater level of ambiguity into the descriptive process.
TABLE OF CONTENTS, FIGURES & TABLES

Appendix 1 Key for the descriptive process model 202
Figure A1 Exemplary descriptive process model illustrating a potential route with and without ambiguities 202

Chapter 2 Appendices 204
Appendix 2.A Information for participants (GAL II) 205
Appendix 2.B Individual graphical descriptions for exemplary listeners (GAL II) 206
Appendix 2.B overview 206
Figures 2.B.1-2.B.6 Listener 1 (programme item 3: stereo) 207
Figs. 2.B.7-2.B.12 Listener 2 (programme item 3: stereo) 207
Figs. 2.B.13-2.B.18 Listener 5 (programme item 3: processed) 208
Figs. 2.B.19-2.B.24 Listener 12 (programme item 1: stereo) 208
Figs. 2.B.25-2.B.30 Listener 6 (programme item 1: stereo) 209
Figs. 2.B.31-2.B.36 Listener 4 (programme item 1: stereo) 209
Figs. 2.B.37-2.B.42 Listener 3 (programme item 1: processed) 210
Figs. 2.B.43-2.B.48 Listener 11 (programme item 1: processed) 210

Appendix 2.C Statistical figures and tables (GAL II) 211
Table 2.C.1 Analysis of variance table for ensemble width with factors of listening location (SEAT), reproduction system (SYSTEM) and listener (LISTENER) 211
Table 2.C.2 Analysis of variance table for ensemble location with factors of listening location (SEAT), reproduction system (SYSTEM) and listener (LISTENER) 211

Appendix 2.D Information for participants (GAL III) 212
Appendix 2.E Individual graphical descriptions for exemplary listeners (GAL III) 213
Appendix 2.E overview 213
Figures 2.E.1-2.E.6 Listener 18 Selected depictions 214
Figs. 2.E.7-2.E.12 Listener 5 Selected depictions 214
Figs. 2.E.13-2.E.18 Listener 3 Selected depictions 215
Figs. 2.E.19-2.E.24 Listener 13 Selected depictions 215
Figs. 2.E.25-2.E.30 Listener 2 Selected depictions 216
Figs. 2.E.31-2.E.35 Listener 9 Selected depictions 216
Figs. 2.E.36-2.E.41 Listener 11 Selected depictions 217
Figs. 2.E.42-2.E.47 Listener 8 Selected depictions 217
Figs. 2.E.48-2.E.53 Listener 1 Selected depictions 218

Appendix 2.F Statistical figures and tables (GAL III) 219
Table 2.F.1 Analysis of variance table for instrument width with factors of programme item (PROGRAM), reproduction system (SYSTEM) and listener (LISTENER) 219
Table 2.F.2 Analysis of variance table for instrument location with factors of listening location (SEAT), programme item (PROGRAM) and reproduction system (SYSTEM) 219
Figure 2.F.1 Means and 95% confidence intervals for instrument width depicted for different programme items 220
Figure 2.F.2 Means and 95% confidence intervals for instrument location depicted from different listening locations & different programme items 220

Chapter 3 Appendices 221
Appendix 3.A Information for graphical elicitation study listeners 222
Appendix 3.B Individual graphical descriptions for exemplary listeners 223
Appendix 3.B overview 223
Figs. 3.B.1-3.B.10 Listener 2 Selected depictions 224
Figs. 3.B.11-3.B.20 Listener 3 Selected depictions 225
Figs. 3.B.21-3.B.30 Listener 9 Selected depictions 226
Figs. 3.B.31-3.B.40 Listener 11 Selected depictions 227
Appendices - Table of contents, figures and tables

Figs. 3.B.41-3.B.50 Listener 13 Selected depictions 228
Figs. 3.B.51-3.B.59 Listener 15 Selected depictions 229
Figs. 3.B.60-3.B.69 Listener 16 Selected depictions 230
Figs. 3.B.70-3.B.79 Listener 17 Selected depictions 231
Figs. 3.B.80-3.B.89 Listener 18 Selected depictions 232
Figs. 3.B.90-3.B.99 Listener 19 Selected depictions 233
Figs. 3.B.100-3.B.109 Listener 20 Selected depictions 234

Appendix 3.C Individual verbal descriptor lists for exemplary listeners 235
Listener 2 235
Listener 3 235
Listener 9 236
Listener 11 236
Listener 13 236
Listener 15 238
Listener 16 238
Listener 17 238
Listener 18 239
Listener 19 239
Listener 20 239

Appendix 3.D Excerpts from transcripts of taped panel discussions 241
Lexicon used in transcription 241
Excerpt 1 Panel II (meeting 5) - discussion about ensemble soundstage 242
Excerpt 2 Panel I (meeting 3) - discussion about ensemble soundstage 243
Excerpt 3 Panel III (meeting 4) - discussion about ensembles 245
Excerpt 4 Panel III (meeting 2) - discussion about height 246
Excerpt 5 Panel II (meeting 2) - discussion about height 247
Excerpt 6 Panel IV (meeting 1) - discussion about area 248
Excerpt 7 Panel I (meeting 2) - discussion about ambience 249
Excerpt 8 Panel I (meeting 3) - discussion about ambience 251
Excerpt 9 Panel II (meeting 1) - discussion about ambience 252
Excerpt 10 Panel II (meeting 2) - discussion about ambience 253

Chapter 4 Appendices 254
Appendix Extracts from the U-GALUI tutorial 255
4.A.1 Figure 4.A.1 Tutorial 1st page. Introduction to the screen 255
Figure 4.A.2 Tutorial 2nd page. Introducing descriptors 255
Figure 4.A.3 Tutorial 3rd page. Moving descriptors 255
Figure 4.A.4 Tutorial 5th page. Changing descriptor size 255
Figure 4.A.5 Tutorial 7th page. Creating descriptors 255
Figure 4.A.6 Tutorial last page. Combining all actions 255

Appendix Instructions for listeners using U-GAL 256
4.A.2

Appendix 4.B Individual listener consistency plots 257
Appendix 4.B overview 257
4.B.1 Agreed use of U-GAL (Listeners 2 and 10) 238
4.B.2 Agreed use of U-GAL (Listeners 1 and 22) 259
4.B.3 Some individuality in U-GAL descriptor use (Listeners 3 and 13) 260
4.B.4 Some individuality in U-GAL descriptor use (Listeners 6 and 21) 261
4.B.5 Individuality in U-GAL use (Listeners 18 and 9) 262
4.B.6 Individuality in U-GAL use (Listeners 11 and 15) 263
4.B.7 Individuality in U-GAL use (Listeners 7 and 16) 264
4.B.8 Individuality in U-GAL use (Listeners 20 and 8) 265
4.B.9 Further examples of visual plots for individual listeners (Listeners 5 and 17) 266

N Ford Doctoral Thesis 199
Appendices - Table of contents, figures and tables

4.B.10 Further examples of visual plots for individual listeners (Listener 19)

Appendix 4.C Overlaid descriptor plots

Appendix 4.C overview

4.C.1 U-GAL Letter Plots
Plots 1-8 Use of letters within FOCAL descriptors
Plots 9-16 Use of letters within AREA descriptors

4.C.2 FOCAL plots (solo instruments only)
Plots 1-4 Use of FOCAL without AREA
Plots 5-8 Use of FOCAL as a localisable centre/core within AREA

4.C.3 AREA and FOS plots (solo instruments only)
Plots 1-4 Use of AREA
Plots 5-8 Use of FOS without AREA

4.C.4 AREA Plots (ensembles by instrument)
Plots 1 – 3 Cello Percussion Voice narrow ensemble
Plots 4 – 6 Voice Cello Percussion wide ensemble
Plots 7 – 9 Cello Percussion Voice right ensemble
Plots 10 – 12 Voice Cello Percussion left ensemble

4.C.5 Plots of ensembles by instrument where FOS was used without AREA
Plots 1 – 3 Cello Percussion Voice narrow ensemble
Plots 4 – 6 Voice Cello Percussion wide ensemble
Plots 7 – 9 Cello Percussion Voice right ensemble
Plots 10 – 12 Voice Cello Percussion left ensemble

4.C.6 Plots of ensembles by instrument where FOCAL was used within an AREA
Plots 1 – 3 Cello Percussion Voice narrow ensemble
Plots 4 – 6 Voice Cello Percussion wide ensemble
Plots 7 – 9 Cello Percussion Voice right ensemble
Plots 10 – 12 Voice Cello Percussion left ensemble

Appendix 4.D Statistical graphs and tables

Table 4.D.1 Analysis of variance table for the variable stimuli, including an estimate of effect size

Figure 4.D.1 Graph illustrating the proportion of the total variability in described location that can be accounted for by the variation in the independent variable stimuli.

Table 4.D.2 Table of non-significant differences from post hoc comparisons of programme material location. (* indicates results were obtained using the Games-Howell and not the Bonferroni procedure)

Figure 4.D.2 Means and 95% CI for listener 13 when describing the position of different stimuli
Figure 4.D.3 Means and 95% CI for listener 1 when describing the position of different stimuli
Figure 4.D.4 Means and 95% CI for listener 22 when describing the position of different stimuli
Figure 4.D.5 Means and 95% CI for listener 15 when describing the position of different stimuli
Figure 4.D.6 Means and 95% CI for listener 17 when describing the position of different stimuli
Figure 4.D.7 Means and 95% CI for listener 18 when describing the position of different stimuli
Figure 4.D.8 Means and 95% CI for listener 9 when describing the position of different stimuli
Figure 4.D.9 Means and 95% CI for listener 6 when describing the position of different stimuli
Figure 4.D.10 Means and 95% CI for listener 5 when describing the position of different stimuli
Figure 4.D.11 Means and 95% CI for listener 10 when describing the position of different stimuli
Appendices - Table of contents, figures and tables

| Figure 4.D.12 | Means and 95% CI for listener 16 when describing the position of different stimuli | 278 |
| Figure 4.D.13 | Means and 95% CI for listener 19 when describing the position of different stimuli | 278 |
| Figure 4.D.14 | Means and 95% CI for listener 8 when describing the position of different stimuli (measured using centre of FOS or AREA descriptor) | 279 |
| Figure 4.D.15 | Means and 95% CI for listener 8 when describing the position of different stimuli (measured using position of letters) | 279 |
| Figure 4.D.16 | Means and 95% CI for existing and naive listeners describing solo instrument sources | 279 |
| Figure 4.D.17 | Means and 95% CI for experienced and inexperienced listeners describing solo instruments | 279 |
| Figure 4.D.18 | Means and 95% CI for the left of centre percussion source as described by all listeners | 280 |
| Figure 4.D.19 | Means and 95% CI for the central percussion source as described by all listeners | 280 |
| Figure 4.D.20 | Means and 95% CI for the left of centre cello source as described by all listeners | 280 |
| Figure 4.D.21 | Means and 95% CI for the central cello source as described by all listeners | 280 |

Appendix 4.E  U-GAL questionnaire responses | 281
Appendix 1 Key for the descriptive process model

The descriptive process model is a means of illustrating the different routes that can be taken between the listeners' auditory spatial experiences and the researcher's comprehension of these experiences. Firstly, the process moves from the listener's experience (Exp₁) to their graphical representation of this experience (R) using either individual (I) or universal (U) descriptors. Secondly, the process progresses to the researcher's comprehension (C) of the listeners' experiences when communicated by means of their graphical representation. The researcher's comprehension can either be representative of the listener's experience (CR) or a mis-comprehension (CM). This determination is based on the presence of problematic ambiguities (any ambiguity which suggests the listener may have had difficulties graphically representing their experiences or one which could prevent the researcher from understanding the listeners' experiences through their graphical representations) in the preceding stages of the descriptive process. As illustrated in figure A1, an ambiguity in the model is illustrated as a dotted line between ensuing stages. A progression which is without ambiguity is denoted by a solid line. Elements of the descriptive process which are not considered by the analysis are represented in grey. Specific elements of the descriptive process are presented in the accompanying table.

Figure A1: Exemplary descriptive process model illustrating a potential route with and without ambiguities

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q</td>
<td>The written instruction provided by the researcher (typically a verbal definition of spatial attributes)</td>
</tr>
<tr>
<td>St</td>
<td>The audio stimulus which gives rise to a listener's experiences</td>
</tr>
<tr>
<td>Exp₁</td>
<td>The listener's experiences that are being represented graphically</td>
</tr>
<tr>
<td>Exp²</td>
<td>The experiences that are not being represented graphically</td>
</tr>
<tr>
<td>R</td>
<td>A listener's representation which describes their experiences with no obvious ambiguity (can be achieved using either individual or universal graphical descriptors)</td>
</tr>
</tbody>
</table>

N Ford
Doctoral Thesis
202
<table>
<thead>
<tr>
<th>Key</th>
<th>The successful comprehension of a listener’s experiences by the researcher</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>Unidentified ambiguities</td>
</tr>
<tr>
<td>IR</td>
<td>Graphical representation in individual language which describes a listener’s experiences. However, unidentified ambiguities are present</td>
</tr>
<tr>
<td>IRQ</td>
<td>Graphical representation in individual language which describes the written instruction provided by the researcher and not the listener’s experiences</td>
</tr>
<tr>
<td>IR1</td>
<td>Graphical representation in individual language which describes a listener’s experiences</td>
</tr>
<tr>
<td>IR2</td>
<td>Graphical representation in individual language which describes a listener’s experiences. However, listener is not responding to the written instruction provided by researcher</td>
</tr>
<tr>
<td>URQ</td>
<td>Graphical representation in universal language which describes the written instruction provided by the researcher and not the listener’s experiences</td>
</tr>
<tr>
<td>UR1</td>
<td>Graphical representation in universal language which describes a listener’s experiences</td>
</tr>
<tr>
<td>UR2</td>
<td>Graphical representation in universal language which describes a listener’s experiences. However, listener is not responding to the written instruction provided by researcher</td>
</tr>
<tr>
<td>UR3</td>
<td>Graphical representation in universal language which does not describe a listener’s experiences</td>
</tr>
<tr>
<td>UR4</td>
<td>Graphical representation in universal language which describes a listener’s experiences. However, listener is not using universal descriptors as agreed.</td>
</tr>
<tr>
<td>UR5</td>
<td>Graphical representation in universal language which does not describe a listener’s experiences. Furthermore, listener is not responding to the written instruction provided by the researcher</td>
</tr>
<tr>
<td>CR</td>
<td>A comprehension by the researcher which is representative of a listener’s experiences</td>
</tr>
<tr>
<td>CR1</td>
<td>A comprehension by the researcher which has the potential for being representative of a listener’s experiences (the use of individual language prohibits a definitive judgement from being made)</td>
</tr>
<tr>
<td>CQ</td>
<td>A mis-comprehension by the researcher resulting from the listener responding to the written instruction and not describing their experiences</td>
</tr>
<tr>
<td>CM</td>
<td>A mis-comprehension by the researcher resulting from unidentified ambiguities in earlier part of the descriptive process</td>
</tr>
<tr>
<td>CM1</td>
<td>A mis-comprehension by the researcher resulting from how the listener has used their individual language</td>
</tr>
<tr>
<td>CM2</td>
<td>A mis-comprehension by the researcher resulting from the listener representing their experiences and not the written instruction</td>
</tr>
<tr>
<td>CM3</td>
<td>A mis-comprehension by the researcher resulting from the listener’s inability to use descriptors from the universal graphical language</td>
</tr>
<tr>
<td>CM4</td>
<td>A mis-comprehension by the researcher resulting from the listener not using descriptors from the universal graphical language as agreed</td>
</tr>
<tr>
<td>CM5</td>
<td>A mis-comprehension by the researcher resulting from the listener’s inability to use descriptors from the universal graphical language. Furthermore, the listener is not responding to the written instruction</td>
</tr>
</tbody>
</table>
CHAPTER 2 APPENDICES
Appendix 2.A Information for participants (GAL II)

Information For Listeners
Please read the following instructions carefully, making sure you fully understand what is required during the investigation. Should you have any questions or queries, please ask!

Introduction
In a moment you will be asked to listen critically for spatial qualities (specified below) of car audio reproductions and respond by DRAWING what you have heard on the response sheets provided. You will be asked to move between three seats within a car and provide responses from each seat.

It is important to note that there are NO CORRECT ANSWERS, instead the investigation is looking for appropriate visual depictions of what YOU have perceived. Therefore, assume what you have heard to be correct and draw this as best as possible.

What you will hear & what qualities to depict
You will be played pieces of music specially recorded for the purpose of this investigation. Each piece consists of three instruments (voice, cello and percussion) positioned within an ensemble.

You are asked to provide the following spatial information on your response sheets:

- **Location of instruments within space**
  Draw the three individual instruments on the response sheet at the location where YOU perceive them to be.

- **Width of ensemble within space**
  Whilst placing the three instruments on the response sheet, think about the width of the ensemble (this is the total width of the three instruments) and draw the ensemble with this width.

Drawing Style
You may use any drawing style you wish to enable you to create an accurate spatial representation of what you have heard. You may label the individual instruments within the image to clarify your depiction as necessary.

Drawing Boundaries
You may draw outside of the car depicted on the response sheet as and when necessary.

Investigation Procedure & Completing the response sheets

- The investigation is split up into 'runs'.
- For a single run you will be given three response sheets.
- You are asked to complete one response sheet for each of the three seating locations within the car.
- The three seats you will be listening from will be pointed out prior to you commencing each run.
- No time limit is placed on your completion of each run.

At the top of each sheet please record your listener number (1 – 12) and the run number.

On the car depicted within the response sheet, mark with a cross the seat you are currently listening from. (There will be three crosses per run (one per seat), indicating the three different seats)
Appendix 2.B Individual graphical descriptions for exemplary listeners (GAL II)

Appendix 2.B overview

This appendix contains graphical descriptions from eight of the 12 participants involved in the second GAL investigation. The six exemplary depictions provided for each of these participants illustrate how the same item of programme material was described from the three listening locations on two different occasions.

Twelve responses are illustrated on each page of the appendix, six each for two different listeners. Pages are organised in two lines with descriptions for one listener occupying the whole of one line. For each line of descriptions, a listener’s responses are arranged so that the two depictions from the driver’s seat occur first, followed by those from the rear-centre listening location and finally the repeated descriptions from the rear-left listening location.

Within the appendix, listeners descriptions are either of programme item 1 or programme item 3. Programme item 1 was a trio ensemble where the percussion occupied a location 20° left of centre, the cello was amplitude panned to the centre of the stereo scene and the voice occupied a location 10° right of centre. In Programme item 3, the same locations within the two-channel recording were employed, but the order of instruments was changed so that the cello was on the left, the voice at the centre, and the percussion to the right of centre.

When listeners were providing these exemplary descriptions, programme item 1 or item 3 were being reproduced using the audio system in its stereo or processed setting. Briefly, the stereo system used six of the seven loudspeakers within the vehicle and generated a phantom centre, whereas the processed setting used all seven loudspeakers and manipulated the two-channel stereo material in order to create a distinct signal at each of these loudspeakers. A note is made of the system setting in operation.

Even though the depictions in appendix 2.B do not show listeners’ responses to the same programme material (listeners did not describe the same stimuli in their repeated runs) it is possible to obtain an indication of each listener’s consistency when representing the same stimulus and to establish whether similarities and differences existed between listeners in terms of their depiction styles.
Appendix 2.C Statistical figures and tables (GAL II)

Table 2.C.1 Analysis of variance table for ensemble width with factors of listening location (SEAT), reproduction system (SYSTEM) and listener (LISTENER)

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig</th>
<th>Partial Eta²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>77895.813</td>
<td>71</td>
<td>1097.124</td>
<td>6.675</td>
<td>.000</td>
<td>.767</td>
</tr>
<tr>
<td>Intercept</td>
<td>653727.522</td>
<td>1</td>
<td>653727.522</td>
<td>3977.312</td>
<td>.000</td>
<td>.985</td>
</tr>
<tr>
<td>SEAT</td>
<td>22447.705</td>
<td>2</td>
<td>1123.853</td>
<td>6.838</td>
<td>.001</td>
<td>.087</td>
</tr>
<tr>
<td>SYSTEM</td>
<td>832.751</td>
<td>1</td>
<td>632.751</td>
<td>3.850</td>
<td>.052</td>
<td>.026</td>
</tr>
<tr>
<td>LISTENER</td>
<td>60890.795</td>
<td>11</td>
<td>5517.900</td>
<td>33.571</td>
<td>.000</td>
<td>.719</td>
</tr>
<tr>
<td>SEAT * SYSTEM</td>
<td>3600.856</td>
<td>2</td>
<td>1600.428</td>
<td>10.954</td>
<td>.000</td>
<td>.132</td>
</tr>
<tr>
<td>SEAT * LISTENER</td>
<td>4506.272</td>
<td>22</td>
<td>204.921</td>
<td>1.247</td>
<td>.218</td>
<td>.160</td>
</tr>
<tr>
<td>SYSTEM * LISTENER</td>
<td>3196.432</td>
<td>11</td>
<td>285.403</td>
<td>1.736</td>
<td>.071</td>
<td>.117</td>
</tr>
<tr>
<td>SEAT * SYSTEM * LISTENER</td>
<td>3070.002</td>
<td>22</td>
<td>139.546</td>
<td>.849</td>
<td>.860</td>
<td>.115</td>
</tr>
<tr>
<td>Error</td>
<td>23668.440</td>
<td>144</td>
<td>164.364</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>755291.775</td>
<td>216</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>101564.253</td>
<td>215</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R Squared = .767 (Adjusted R Squared = .652)

Table 2.C.2 Analysis of variance table for ensemble location with factors of listening location (SEAT), reproduction system (SYSTEM) and listener (LISTENER)

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig</th>
<th>Partial Eta²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>64363.038</td>
<td>71</td>
<td>906.530</td>
<td>4.966</td>
<td>.000</td>
<td>.710</td>
</tr>
<tr>
<td>Intercept</td>
<td>61665.342</td>
<td>1</td>
<td>61665.342</td>
<td>337.821</td>
<td>.000</td>
<td>.701</td>
</tr>
<tr>
<td>SEAT</td>
<td>14977.393</td>
<td>2</td>
<td>7488.686</td>
<td>41.025</td>
<td>.000</td>
<td>.363</td>
</tr>
<tr>
<td>SYSTEM</td>
<td>12691.767</td>
<td>1</td>
<td>12691.767</td>
<td>69.529</td>
<td>.000</td>
<td>.326</td>
</tr>
<tr>
<td>LISTENER</td>
<td>11509.007</td>
<td>11</td>
<td>1046.273</td>
<td>5.732</td>
<td>.000</td>
<td>.305</td>
</tr>
<tr>
<td>SEAT * SYSTEM</td>
<td>8863.767</td>
<td>2</td>
<td>4431.884</td>
<td>24.279</td>
<td>.000</td>
<td>.252</td>
</tr>
<tr>
<td>SEAT * LISTENER</td>
<td>9918.981</td>
<td>22</td>
<td>450.893</td>
<td>2.470</td>
<td>.001</td>
<td>.274</td>
</tr>
<tr>
<td>SYSTEM * LISTENER</td>
<td>3556.716</td>
<td>11</td>
<td>323.338</td>
<td>1.771</td>
<td>.064</td>
<td>.119</td>
</tr>
<tr>
<td>SEAT * SYSTEM * LISTENER</td>
<td>2845.988</td>
<td>22</td>
<td>129.383</td>
<td>.709</td>
<td>.825</td>
<td>.098</td>
</tr>
<tr>
<td>Error</td>
<td>26285.570</td>
<td>144</td>
<td>182.539</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>152314.550</td>
<td>216</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>90649.208</td>
<td>215</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R Squared = .710 (Adjusted R Squared = .567)
Appendix 2.D Information for participants (GAL III)

Information For Listeners
Please read the following instructions carefully, making sure you fully understand what is required during the investigation. Should you have any questions or queries, please ask!

Introduction
In a moment you will be asked to listen critically for spatial qualities (specified below) of audio reproductions and respond by DRAWING what you have heard on the response sheets provided. You will be asked to move between two seats within a vehicle and provide responses from each seat.

It is important to note that there are NO CORRECT ANSWERS, instead the investigation is looking for appropriate visual depictions of what YOU have perceived. Therefore, assume what you have heard to be correct and draw this as best as possible.

What you will hear & what qualities to depict
You will be played pieces of music specially recorded for the purpose of this investigation. Each piece of music consists of an instrument from the following list: male voice, female voice, cello and percussion.

You are asked to provide the following information about what you hear on your response sheets:

Location of instrument within space
Draw the individual instrument on the response sheet at the location where YOU perceive it to be.

Width of instrument within space
Whilst placing the instrument on the response sheet, think about how wide the instrument appears to be, and draw the instrument with this width.

Drawing Style
You may use any drawing style to enable you to create an accurate spatial representation of what you have heard. You may label the individual instruments within the image to clarify your depiction as necessary.

Drawing Boundaries
You may draw outside of the car depicted on the response sheet as and when necessary.

Investigation Procedure & Completing the response sheets
- The investigation is split up into 'runs'.
- For a single run you will be given two response sheets.
- You are asked to complete one response sheet for each of the two listening locations.
- The two seats you will be listening from will be pointed out to you prior to commencing each run.
- no time limit is placed on your completion of each run.

On each response sheet, mark with a cross the seat you are currently listening from.
Appendix 2.E Individual graphical descriptions for exemplary listeners (GAL III)

Appendix 2.E overview

This appendix contains graphical descriptions from nine of the 31 participants involved in the third GAL investigation. Exemplary depictions illustrate how each listener chose to describe a selection of the programme material when reproduced over the various audio system settings.

Figures 2.E.1-2.E.6, 2.E.7-2.E.12 and 2.E.13-2.E.18 depict three listeners’ responses to the programme items reproduced over different system settings. Figures show how listeners described the cello and percussion when reproduced using the optimal audio system (with full bandwidth centre channel) setting, the same instruments reproduced using the reduced bandwidth centre and finally a phantom centre. Exemplary descriptions for listeners 18, 5 and 3 illustrate how each described differences when the material was reproduced using the different settings; descriptors getting progressively smaller and more skewed as the system used a less optimal setting. Listeners’ responses were not ambiguous to the researcher as they matched the researcher’s own knowledge of the situation.

Figures 2.E.19 – 2.E.24 illustrate how one listener, (listener 13) demonstrated consistency when repeatedly describing the same instrument; the cello reproduced using the optimal and phantom system settings. Moreover, the listener described differences between audio system settings, with both cello and percussion stimuli described as narrower when the phantom rather than the optimal setting was used. Again, this listener’s responses were not identified as ambiguous.

Listeners 2 and 9 (figures 2.E.25-2.E.30 and figures 2.E.31-2.E.35 respectively) were amongst those listeners whose descriptive styles were not so straightforward to understand. Listener 2 was inclined to split their depictions of a single instrument and place these at various loudspeakers around the vehicle. Listener 9 took time to develop their individual language, starting with more complex descriptors as illustrated in figures 2.E.33-2.E.35 and ending with a far more simple descriptive style (figure 2.E.31 and 2.E.32). Although more difficult, these descriptors could still be understood by the researcher.

The graphical representations of listeners 11 and 8 (figures 2.E.36-2.E.41 and figures 2.E.42-2.E.47 respectively) were ambiguous. For these, and other listeners, descriptors were on the opposite side of the vehicle to the listening location and, for listener 11, descriptors were positioned to the rear of the vehicle. No obvious connection could be made between these representations and the stimuli, and there did not appear to be a recognisable pattern. Although the depictions could have reflected the listeners’ experiences, ambiguities prevented the researcher from understanding what was being described.

Figures 2.E.48 – 2.E.53 illustrate how listener 1’s descriptive style was not consistent. Furthermore descriptors covered the front of the vehicle regardless of audio stimulus. Again these representations could have reflected the listeners’ experiences, but the inclusion of various elements within each representation suggested that something other than width and location gave rise to these responses.
Figures 2.E.25 - 2.E.30 Listener 2 Selected depictions

- Figure 2.E.25: Cello (Reduced)
- Figure 2.E.26: Cello (Reduced Passenger)
- Figure 2.E.27: Cello (Phantom)
- Figure 2.E.28: Cello (Phantom Passenger)
- Figure 2.E.29: Percussion (Phantom)
- Figure 2.E.30: Percussion (Phantom Passenger)

Figures 2.E.31 - 2.E.35 Listener 9 Selected depictions

- Figure 2.E.31: Cello (Optimal)
- Figure 2.E.32: Cello (Reduced)
- Figure 2.E.33: Cello (Phantom)
- Figure 2.E.34: Cello (Phantom Passenger)
- Figure 2.E.35: Percussion (Optimal)
Appendix 2.F Statistical figures and tables (GAL III)

Table 2.F.1 Analysis of variance table for instrument width with factors of programme item (PROGRAM), reproduction system (SYSTEM) and listener (LISTENER)

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>418006.070</td>
<td>371</td>
<td>1126.701</td>
<td>7.424</td>
<td>.000</td>
<td>.881</td>
</tr>
<tr>
<td>Intercept</td>
<td>1403719.011</td>
<td>1</td>
<td>1403719.011</td>
<td>924.903</td>
<td>.000</td>
<td>.961</td>
</tr>
<tr>
<td>LISTENER</td>
<td>268046.947</td>
<td>30</td>
<td>9601.565</td>
<td>63.269</td>
<td>.000</td>
<td>.836</td>
</tr>
<tr>
<td>PROGRAM</td>
<td>21922.780</td>
<td>3</td>
<td>7307.933</td>
<td>48.153</td>
<td>.000</td>
<td>.890</td>
</tr>
<tr>
<td>SYSTEM</td>
<td>17913.244</td>
<td>2</td>
<td>8956.622</td>
<td>59.019</td>
<td>.000</td>
<td>.841</td>
</tr>
<tr>
<td>LISTENER * PROGRAM</td>
<td>27804.723</td>
<td>90</td>
<td>308.941</td>
<td>2.036</td>
<td>.000</td>
<td>.330</td>
</tr>
<tr>
<td>LISTENER * SYSTEM</td>
<td>23467.218</td>
<td>60</td>
<td>391.120</td>
<td>2.577</td>
<td>.000</td>
<td>.294</td>
</tr>
<tr>
<td>PROGRAM * SYSTEM</td>
<td>5114.320</td>
<td>6</td>
<td>852.387</td>
<td>5.617</td>
<td>.000</td>
<td>.833</td>
</tr>
<tr>
<td>LISTENER * PROGRAM * SYSTEM</td>
<td>33736.838</td>
<td>180</td>
<td>167.427</td>
<td>1.235</td>
<td>.047</td>
<td>.374</td>
</tr>
<tr>
<td>Error</td>
<td>58454.078</td>
<td>372</td>
<td>151.758</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1878179.159</td>
<td>744</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>474460.148</td>
<td>743</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R Squared = .881 (Adjusted R Squared = .762)

Table 2.F.2 Analysis of variance table for instrument location with factors of listening location (SEAT), programme item (PROGRAM) and reproduction system (SYSTEM)

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>1481649.908</td>
<td>23</td>
<td>64419.581</td>
<td>82.966</td>
<td>.000</td>
<td>.726</td>
</tr>
<tr>
<td>Intercept</td>
<td>870.737</td>
<td>1</td>
<td>870.737</td>
<td>1.121</td>
<td>.290</td>
<td>.002</td>
</tr>
<tr>
<td>PROGRAM</td>
<td>1741.343</td>
<td>3</td>
<td>580.448</td>
<td>.748</td>
<td>.524</td>
<td>.003</td>
</tr>
<tr>
<td>SEAT</td>
<td>1312071.827</td>
<td>1</td>
<td>1312071.827</td>
<td>1889.817</td>
<td>.000</td>
<td>.701</td>
</tr>
<tr>
<td>SYSTEM</td>
<td>2184.317</td>
<td>2</td>
<td>1092.159</td>
<td>1.407</td>
<td>.246</td>
<td>.004</td>
</tr>
<tr>
<td>PROGRAM * SEAT</td>
<td>47974.335</td>
<td>3</td>
<td>15991.445</td>
<td>20.595</td>
<td>.000</td>
<td>.079</td>
</tr>
<tr>
<td>PROGRAM * SYSTEM</td>
<td>817.026</td>
<td>6</td>
<td>136.171</td>
<td>.715</td>
<td>.983</td>
<td>.001</td>
</tr>
<tr>
<td>SEAT * SYSTEM</td>
<td>111699.071</td>
<td>2</td>
<td>55649.535</td>
<td>71.929</td>
<td>.000</td>
<td>.167</td>
</tr>
<tr>
<td>PROGRAM * SEAT * SYSTEM</td>
<td>5161.988</td>
<td>8</td>
<td>860.331</td>
<td>1.108</td>
<td>.356</td>
<td>.009</td>
</tr>
<tr>
<td>Error</td>
<td>559049.849</td>
<td>720</td>
<td>776.458</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2041570.494</td>
<td>744</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>2040699.756</td>
<td>743</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R Squared = .726 (Adjusted R Squared = .717)
Figure 2.F.1 Means and 95% confidence intervals for instrument width depicted for different programme items.

Figure 2.F.2 Means and 95% confidence intervals for instrument location depicted from different listening locations and different programme items.
Appendix 3.A Information for graphical elicitation study participants

Information for Listeners
Please read the following instructions carefully, making sure you fully understand what is required during the investigation. Should you have any questions or queries, please ask!

Introduction
In a moment you will be asked to listen critically for spatial qualities (specified below) of audio reproductions and respond by DRAWING what you have heard on the response sheets provided.

It is important to note that there are NO CORRECT ANSWERS, instead the investigation is looking for appropriate visual depictions of what YOU have perceived. Therefore, assume what you have heard to be correct and draw this as best as possible.

What you will hear & what qualities to depict
You will be played pieces of music specially recorded for the purpose of this investigation. Each piece of music consists of either solo drums or solo cello, or a combination of drums, cello and female voice playing simultaneously.

You are asked to provide the following information about what you hear on your response sheets:

- **Location of instrument(s) within space**
  Draw the instruments on the response sheet at the location where YOU perceive them to be.

- **Width of individual instrument(s) within space**
  Whilst placing the two solo instruments (drum & cello) on the response sheet think about how wide the instrument appears to be and draw the instrument with this width.

- **Width of ensemble within space**
  Whilst placing the ‘ensemble’ instruments (tracks with simultaneous cello, drums & female voice) on the response sheet, think about how wide the ‘ensemble’ of these instruments appears to be, and draw the ensemble with this width.
  NB: you do not have to provide a width for the individual instruments within the ensemble.

Drawing Style
You may use any drawing style to enable you to create an accurate spatial representation of what you have heard. You may label the individual instruments within the image to clarify your depiction as necessary.

Drawing Boundaries
You may draw outside of the car depicted on the response sheet as and when necessary.
Appendix 3.B Individual graphical descriptions for exemplary listeners

Appendix 3.B overview

Appendix 3.B contains a selection of graphical depictions from individual listeners involved in the clarification investigation.

Exemplary descriptions are presented in numerical order (from listener 1 to listener 20). The depictions of one listener occupy one page. For the majority of listeners, representations are illustrated for the listeners' auditory spatial experiences of the same items of programme material, these items are:

- **Percussion Left**: The percussion extract amplitude panned fully left in the two-channel stereo scene
- **Percussion Central**: The percussion amplitude panned to the centre of the stereo scene
- **Perc Central (Rpt)**: A repeated depiction of the same percussion central stimulus as perceived
- **Cello Left**: The cello extract amplitude panned fully left in the two-channel stereo scene
- **Cello Central**: The cello extract amplitude panned to the centre of the stereo scene
- **Cello Central (Wide)**: The same central cello stimulus, but with artificial stereo reverberation added
- **CPV Narrow**: Trio ensemble of cello, percussion and voice closely spaced around the centre of the stereo scene in the recording
- **VCP Wide**: Trio ensemble with the voice panned fully left, cello central and percussion fully right in the stereo scene
- **VCP Left**: Trio ensemble with the voice panned fully left, cello centre-left and percussion at the centre of the stereo scene
- **CPV Right**: Trio ensemble where the cello was amplitude panned to the centre of the stereo scene, the percussion mid-way between the centre and full right and the voice panned fully right.

Not all listeners completed the same repeats. One listener (listener 15) did not repeatedly describe his auditory spatial experiences of any of the stimuli included in this appendix. Consequently no repeats are presented for this listener. Other listeners were required to repeatedly describe the **CPV Right** ensemble instead of the **Percussion Central**. For these listeners, their repeated descriptions of the auditory spatial scene invoked by this ensemble are included in the appendix.
Figures 3.B.1 – 3.B.10 Listener 2 Selected depictions
3.B.1 Percussion Left 3.B.2 Percussion Central 3.B.3 Cello Left 3.B.4 Cello Central
3.B.5 Cello Central (Wide)
3.B.10 CPV Right (Rpt)
Figures 3.B.41 - 3.B.50 Listener 13 Selected depictions
Figures 3.B.51 – 3.B.59 | Listener 15 Selected depictions

3.B.51 Percussion Left
3.B.52 Cello Left
3.B.53 Cello Central
3.B.54 Percussion Central
3.B.55 Cello Central (Wide)
3.B.56 CPV Narrow
3.B.57 CPV Wide
3.B.58 VCP Left
3.B.59 CPV Right
Figures 3.B.70 - 3.B.79 Selected depictions

3.B.70 Percussion Left
3.B.71 Percussion Central
3.B.72 Perc Central (Rpt)
3.B.73 Cello Left
3.B.74 Cello Central
3.B.75 Cello Central (W ide)
3.B.76 CPV Narrow
3.B.77 VCP Wide
3.B.78 VCP Left
3.B.79 CPV Right

3.B.73 Cello Left
Figures 3.B.100 – 3.B.109 Listener 20 Selected depictions


Appendix 3.C Individual verbal descriptor lists for exemplary listeners

LISTENER 2
Position
Width
Big (bigger)
Size of the Image (large) (small)
Image shift / Image move
In front of me / right / extreme right / extreme left / centre
Very small
Coming from the same place
Size of the circle (large)
Height
Further up in the car
Ambience (no ambient information)
Behind me (more behind, less behind)
Right around
Image in front - instrument in front (main cue)
Size of the instrument behind (envelopment)
Normal sizes (no ambience)
Point source (small)
Bigger than point source
Condensed
Depth (instrument further away)
Instrument(s) in front of other Instrument(s)
Spread out
Overlapping circles

LISTENER 3
Specific sound position (comes from specific area)
Area sound comes from
Couldn’t work out where sound coming from (difficult)
Wide / narrow (narrow area) (same width)
Broodness (broader) (not very broad sound)
Size of Instrument
Volume
Middle = louder
Sound from centre / sound towards centre
Sound from front of car
Down / up / high / below the vent
Attention attracted towards speaker
Sound coming towards me
Specific vs. broad
Sound directly in front of me / coming directly at me
Behind
Specific narrow sound (small)
Sound coming from above

Listener 2: Ensembles
Tight together instruments
Definite places
Detect where difference places in ensemble are
Easy to pick out (instruments)
Width (has width)
Small stage (light) (doesn’t have width)
Envelopment
Close together
Image shift
Centre / side / well over into side of car
One circle / overlapping circle
Width of soundstage
Size of soundstage
Well defined (spaced) instruments within space
Joined / not joined
Broodness of individual source
Positions
Bigger
Wider / width (narrow)
Small soundfield
Can pick out positions
Across
Big image / big soundstage
Phase (out of phase)
Fairly small (tight together)

Listener 3: Ensembles
Area
Broad (quite)
Specific (very narrow) (quite narrow)
Sounds coming from same place / instruments all in same area
Proximity
Coming from a broad area
Instrument(s) coming over the top
Can’t put finger on where instrument is coming from
Instrument (not) coming towards me
Individual width
Easily picked up (quite narrow)
Wider (broader)
Instruments behind each other
Instrument further away vs. instrument closer
Instrument towards the front of the car
Ensemble wide vs. in specific area (narrow)
Instrument isolated
Nothing from centre of vehicle
<table>
<thead>
<tr>
<th>Listener 9: Ensembles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Width across boxes (ensemble size)</td>
</tr>
<tr>
<td>Strong image</td>
</tr>
<tr>
<td>Sound covers a wider area</td>
</tr>
<tr>
<td>Size of source</td>
</tr>
<tr>
<td>Instrument covers a bigger area</td>
</tr>
<tr>
<td>Very big image</td>
</tr>
<tr>
<td>Wide images</td>
</tr>
<tr>
<td>Wide source</td>
</tr>
<tr>
<td>Instrument spreads around</td>
</tr>
<tr>
<td>Image moves around an area</td>
</tr>
<tr>
<td>Sense of spaciousness</td>
</tr>
<tr>
<td>Not very much width (localised sound)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Listener 11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sound in corner (steered into corner)</td>
</tr>
<tr>
<td>Sound comes across into centre</td>
</tr>
<tr>
<td>Out of centre speaker / definitely from centre</td>
</tr>
<tr>
<td>Locating / located / locality / location</td>
</tr>
<tr>
<td>Width (wider) (more width)</td>
</tr>
<tr>
<td>Louder (loudness)</td>
</tr>
<tr>
<td>Soundstage (wider)</td>
</tr>
<tr>
<td>Sound from door speaker</td>
</tr>
<tr>
<td>Direction</td>
</tr>
<tr>
<td>Drop in image height (position)</td>
</tr>
<tr>
<td>Instrument up / Instrument lower / different level</td>
</tr>
<tr>
<td>Greater distance</td>
</tr>
<tr>
<td>Broader soundstage / larger soundstage</td>
</tr>
</tbody>
</table>
Clarifying individual listeners' descriptions and developing U-GAL

Height of image (low height) (more height)
Volume (Amplitude)
Volume (Area)
Directional sound
Corner image
Narrow(er) image
Scattering
Specific source pulls the image
Cone of sound
Depth (how much sound outside of car) (live sound)
Poor depth (sound that hugs the space)
Depth towards me (from where the sound is coming from to where I am sitting)
Sound to rear / sound coming from behind
Reverberation (something else to sound)
Sound further away
Skew
Less information from centre fill
Hard to locate vs. easy to locate
Sound fills the space
Sound coming from all around me
Nothing from the rear
Resonances
Information from left instrument panel / from door
Confusing for locality
Clear
Precise

Listener 11: Ensembles
Instruments come from same location
Concise unit / concise ensemble
Distinct location for instruments/ difference in location
Width (ensemble narrower (wide) / broader image)
Depth of image visualise more than thin line of sound
Instrument at back of sound / central / at front
Instruments staggered in different locations (depth)
Front - back width
Precise (narrower image) (precise ensemble)
Small image (small depth of image)
Soundstage
Width (wider)
Stage presence
Everything to left
Ensemble hugs the boundaries of the car
Doesn't give me an impression of involvement
Not involved / more involvement (with soundstage & sound)
Ambience (overall) (links to depth)
I'm here, they are there
More information coming towards me

Instrument skewed by another instrument
Centre fill (more information from centre fill)
Definite front - back image
Split in where ensemble image is (Left - right split)
Mixed up (coming from same soundstage - not separated)
Separate / separation
Instruments coming from the same location
Soundstage not wide
Easily localisable
Gap in centre
Instruments linked
More localised image (narrower)
Locality more precise (direction) vs. merging
Gap in image
More information coming from centre
More Front - rear location

Listener 13
Volume (the way the sound comes out)
Soft vs. loud (volume)
Striking
Powerful
Clarity (sharp)
Density of volume (strength)
Where sound is coming from
Sound from centre / from my side / sound from front
Area
Sound comes out both sides (projected both sides)
Area of volume
Spread
Surround (full)
Instruments further down (height)
Projection (where sound is projected to)
Wider spread
Loudness / volume
Where sound is (main)
Sound all around
Coming from (out of) same area / different area /
projected area
Wider area
Instrument comes out to you
Wider area (louder)
Projecting out from corner /projected over wider area
Sound goes further
Area volume spreads to
Sound covered the whole of the area

N Ford
Doctoral Thesis
Instrument blending
Hard to distinguish instruments

LISTENER 15
Specific area (only in the one area)
Generally in this area
(Right) At the back
Only in the back / only in the front
Sound came around
Sound behind me
Sound comes out
Sound just around my head
Not coming up to me
Just one small area vs. bigger area
Fills whole width of the car
Fills whole of the back
Fills (covers) all of the front of the car
Filling the car vs. specific area in car
Fills the whole of the area (not specific)
Sound is everywhere
Sound covers all
Spreading all around
Spread further backwards / right across
Spread (took up a bigger space)
Sound predominately (mainly) in one area
Instrument is there / instrument could be anywhere from here
Couldn’t say where instrument was
Specific place where instrument was being played
Precise (one little location)
Bigger
Loud
Wider (front - back) (fairly wide) (wider)
Width (left - right)
Deep (quite a big area)
Spikes across

LISTENER 16: Ensemble
Narrow / wider
Angle
Width
Taking a back seat to other instruments
Directional
Instrument swamped
Instrument in front
Curved sound

LISTENER 17
Intense/intensity of field (smaller)
How much sound is around / coming from all around
How far sound travels
Enveloping (coming right around / out towards you)
Sound coming from one area / fills an area
Source (where the sound is coming from)
Sound wrapping around listener
Size of field
Listening from a distance (coming to you) vs. sitting amongst what is going on around
Sound coming from certain direction
Wall of sound
Sound coming across
Sound coming from point vs. filling up more of space
Sound around (enveloping atmosphere) vs. at one point (a point source)
Sound coming from one place & broadening out vs. coming from one area and staying there
Listening in large room (music moves out through whole space) vs. in a small room
Sound coming from side vs. coming from one point
Centre fill sound
Enveloping the front of the car
Hard to pick out where source was coming from
Enveloped by sound / sound coming from all area
Amongst sound vs. hearing it from a distance
Two point sources but nothing wrapping you up
Depth of field
Coming from area (enveloped)
Field wrapping around
Didn't feel part of what was going on / sound not coming around listening position
Source from two directions (corner to corner)

Listener 17: Ensembles
Sounds coming from different places in vehicle
Taking part
Not sitting amongst sound vs. sitting amongst sound
(Greater) depth of field
Envelopment (feel of enveloping listener) (involved)
All instruments coming from one source
Ensemble wraps a space
Weaker area
Part of the sound vs. listening to it
Sound predominates vs. at distance
Instrument next to you vs. going on at a distance
Instrument(s) within the field / outside the field
Instrument(s) not part of the main sound
Three distinct sounds
All instruments of equal intensity
Pick out where all instruments were coming from / where sound was going
Listening to sound vs. being involved
Instruments from different areas
Instrument on periphery - not part of main ensemble
One sound vs. three sounds (in different areas)
Listening at distance vs. sitting amongst instruments

LISTENER 18
Place
Width (wide) (wider)
Direction of the width
Localised (highly)
Tight image (tighter)
Point image
Bigger
Epicentre
Has central core
Out of the speakers
Hard fixed local image
Not good image / less focused image
Closer to me
In front of
More forward feel

Listener 18: Ensemble
Width / depth (depending on where you are sitting)
Unfocused (fairly)
Tight
Instruments come from the same place
Fairy narrow
Placed in front
Further away
Depth with respect to other instrument
Small distance
Main direction
Clearer
Height
Clear separation between the instruments
3-D difference

LISTENER 19
Width
Focal point
Area of the sound
Pocket of sound
Envelope of sound (bigger) (tighter)
Tight area
Tight band / Precise band
Acute
Spread across
Sound fills an area
Coming from everywhere
No focal point
Bathed in sound
Sound in front of me / (directly) behind me
More rear speaker
Sound on bonnet / in the corner
Couldn’t pinpoint sound
Quicker to discern / Couldn’t discern
Difficult to pick out where it was coming from

Listener 19: Ensemble
Width
Area
Breadth
Acute
Band
Saturation
Pocket of sound
Fills front of the car
Instrument fills the area (Fills most of the area)
Wide range
Wider band
Narrow area
Width of all instruments the same
Most focused vs. widest
More specific areas vs. lapping over
Sound lapping across (sound overlapping)
Focal point
Focused (more or less)
Distinct sounds
Localised areas
Easier to pick out
Specific area sound is coming from (no overlap)
Could pick out definite sounds
Mixed up sound / sound confusing (not focused)
Instrument difficult to pinpoint
Harder to pick out sound (because sound overlaps)
Can’t distinguish banding
Sound merged
Blending of sound
Sound at centre of dashboard
Instrument sat next to me
Instruments coming from exactly same place (area)
Sound coming from central area
Centre fill dead
Nothing in front of me / Sound right in front of me
Sound straight (right) across me / not across car

Instrument moves around
Sound further back / not as far forward
Area of sound coming towards me
Sound comes into vehicle

LISTENER 20
Shifted
Wider / narrower
Centre of the sound
Ambience of the sound (no ambience)
All centre channel
Image of sound narrow
Equal
Go around me as a listener
Extension to rear of front
Position (uncertain of position)
Difficult to pinpoint
Difficult to locate (simple to locate)
Width is all over the front
Different location
Fill
Centred towards a source (mid point of source)
Wider than a point
Scaling
Level

Listener 20: Ensemble
Position (same)
Difficulty in positioning instrument
Ambience
Clear
Less ambience / more ambience
More difficult to get exact location
Down (height)
Moves back
All towards the centre
Wider
Harder to locate
Appendix 3.D Excerpts from transcripts of taped panel discussions

Lexicon used in transcription

The transcription symbols used in the included excerpts are taken and adapted from the transcript conventions referred to by Silverman (1993). Symbols are as follows:

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Example</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>2: R:</td>
<td>2: 1, =</td>
<td>Numbers identify the panelist currently speaking. 'R' indicates the panel leader is speaking</td>
</tr>
<tr>
<td>=</td>
<td>2: 1, =</td>
<td>Equal signs at the end of one line and beginning of subsequent line indicate no gap between the two lines</td>
</tr>
<tr>
<td>[</td>
<td>14: = or is it like this</td>
<td>Left brackets indicate the point at which a current panelists talk is overlapped by another's talk</td>
</tr>
<tr>
<td>( )</td>
<td>14: Yeees (.) the only thing,</td>
<td>Indicates a pronounced gap between the panelists words</td>
</tr>
<tr>
<td>.hhhh</td>
<td>17: hhhh I don't think</td>
<td>A row of h's prefixed by a dot indicates an 'in' breath, without a dot indicates an 'out' breath. The number of 'h's' indicates the length of the in or out breath</td>
</tr>
<tr>
<td>( )</td>
<td>4: ( ) you'd have arrows</td>
<td>Empty ( ) indicates an inability (by the researcher) to hear what was being said</td>
</tr>
<tr>
<td>( ( ))</td>
<td>5: this one ((area bubble))</td>
<td>(( )) contain researcher's descriptions rather than transcription</td>
</tr>
<tr>
<td>naturalness</td>
<td>11: um, naturalness</td>
<td>Italics indicate a stress on a particular word</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td>Ellipses between text indicates a break in the transcription</td>
</tr>
</tbody>
</table>
Excerpt 1 Panel II (Meeting 5) - discussion about ensemble soundstage

Do you, do you understand what we mean by soundstage?

Yes

Right

Describe it

I would describe the soundstage of an ensemble as the area filled by the complete range of instruments.

Fine, yes

Or the width or whatever

Yes, that's right and is that what it means to you?

Yes, that's fine.

That's like separate from this, then is the soundstage the whole car?

I, 

Or is it like this? I dunno

Well it could be, it's possible

Or it, or it could be that the soundstage is this and there is another instrument over here?

So that is a separate then, there are two soundstages is there?

That would be, that would be a point source

That would be, yes

See that would be a point source there then, wouldn't it?

But

Alright well say this, say there are, there are two things here, instruments here

But, there is no reason why you couldn't have the soundstage do that (indicates a soundstage with two locations)

Can you, I think, could you have two soundstages in one?

Yes, why not, yes

Yes
Excerpt 2 Panel I (Meeting 3) - discussion about ensemble soundstage

R: Is there anything else you need for ensembles?
11: (.) Soundstage
R: Yes, a soundstage, would you indicate (.) soundstage in some way?
11: (.) Then, well you need to look at the um, (.) the width and depth of the soundstage
4: ( ) you’d have arrows across and their location point, with an indication of what you heard =
20: [yes
5: what I found was, those three in there (instruments), they were all close together (looking at response for a particular ensemble) and a lot of them overlapped each other, so how are we going to stop them being, having the confusion of (5 is looking through the pile of graphical responses as he speaks) (.) are they here?
R: Yes they are there (referring to 5’s response sheets), they are there somewhere
5: Something like that (referring to another ensemble source which overlapped), how are we going to (.) separate them out?
4: You don’t want to, do you? You want to indicate it was a =
5: = ( ) That’s very confusing and I had to explain to ((the second researcher)) what I meant by that
R: What do you mean by that?
5: Well he thought the ‘v’ (female speech source) was in this one ((area ‘bubble’)), when it was actually in this one ((different area ‘bubble’))
R: Right (.) [Yes
5: And I know, I know we are not drawing bubbles, we are drawing lines, but lines [ ( )
11: You could, you could have different colours for each (.) for each [instrument
5: You could yes, yes
20: If uh =
11: = ( ) worried about confusion
4: Different coloured arrows? or
11: And letters?
20: Yes
5: Or loudness as well? We are having different colours for loudness are we not!
((general laughter))
11: You could have boldness, the boldness. How (.) is it a thin line, or =
R: Ah yes, cause of course loudness is going to come in for your ensembles
20: Oh yes
R: So you could have different colours representing different instruments, now will each individual instrument (.) could this (.) I’m going back to what ‘S’ has got here and his individual letters indicating the source (.) of the sound, would you put these little markers ((graphical indication of the presence of a ‘main’ part to the sound)) where you thought the source was?
20: Yes
11: Yes
R: Now what would happen if you didn’t think it had a source, cause there are some that you can’t find, you can’t identify the source of the sound, you just know it is there
4: You mean like a location just, a larger location then, rather than a pinpoint location
11: Then, then you could leave the source out
R: Just have a coloured line?
11: Just have a coloured line yes
4: If you can’t localise on it, then you don’t put it in
11: Don’t put it in yes
4: Rather than guessing
R: ((briefly recaps what has occurred up until now, then continues)) The soundstage, we need to, is it important then to do the whole soundstage?

11: I think so, because once you've done the individuals, you then say that (.) collectively they set up a soundstage so you can visualise where the ensemble is, or the band is, you've got width and depth and height and ambience, everything else associates with that soundstage.

4: Are we allowed to, are we allowing it to overlap as well, run into each other?

11: Yes (.) the individual ones like '5's got. '5' had, umm individuals overlapping = ('individual' refers to individual sources)

4: = we all have that I think

11: Its, its total width and depth would be the soundstage for that [ensemble say

4: uhuh

20: If we have the possibility to have a mono signal or whatever of a person speaking at the same time as a band plays in the left corner.

11: Yes, yes, so then you'd, [that

20: so what is the soundstage?

(R interrupted by someone entering the room)

11: Well shall we just finish off what we were just talking about? Um, does everybody agree on soundstage? '20' just said, you could have an ensemble in one corner and somebody speaking in the other corner, then (.) you potentially could have a hole in the centre between the two (.) so where they don't overlap there are two distinct soundstages. So you'd have a soundstage on the left say, where you'd got an ensemble and a single source coming from the right.

4: Yes, so we only allow it to overlap if it needs to but we don't want everything to overlap.

11: No, we don't fill in the space, if there is no sound coming from that =

20: = no, no ((agreeing))

R: If a space exists we are going to allow them to put a space in? =

4: = yes, have a space, yes

11: Yes

(R recaps what has just been discussed and then asks how they are going to deal with the 'space'))

11: It is down to location, if you see that there is somebody speaking in the right and he's got his own width and depth, and there is an ensemble on the left which individually they have got a soundstage with width and depth, then you treat it as two sources, from an ensemble (.) or however else you'd like to describe it as

(Another interruption, session closes)
Excerpt 3 Panel III (Meeting 4) - discussion about ensembles

17: When I did it, I considered the thing ((the ensemble)) as a whole really
R: Right, rather than the individual instruments?
17: Yes, but if I could hear instruments coming from different sources, I'd just put where they were coming from
R: Ok
17: And if (. ) two combined, came from the same source, well then mark it as the same source and then consider the group as a whole
10: Well I did consider each individual instrument sound (. ) I’d sort of pinpoint the cello and then the drums and then the voice (. ) but you didn’t do it like that?
17: No
10: You just sort of said; there it is?
R: So it may be that you don’t necessarily need individual instrument width and depth when dealing with ensembles, but you have an individual instrument origin?
17: Maybe (. ) if you look at what ‘8’ has done there, he’s highlighted individual ones, but then you could almost group that together as a whole sort of thing, couldn’t you?
8: You wouldn’t extract much information if it was just a big circle, would you?
17: No
8: Your width (. ) your area would be (. ) effectively gone then wouldn’t it? for each instrument (. ) your width and your depth.
17: I think, I think you just listen to the ensemble and if you pick up several sounds coming from one place and it all sounds like a combined integrated feel you just sketch it as one field (. ). Say you’ve got, from my point of view, say you’ve got like a voice, cello and a drum (. ) and the drum and the cello are combined in the front here ((drawing on a response sheet as he talks)) and you hear that as one group sound. But you hear (. ) a voice coming from somewhere else and you do two fields, one being the combined of the one (. ) and one being the other. If you hear them all from the same place then you do one field (. ) you just need to [highlight which one is which
8: Yes, I agree with that, yes
17: Because sometimes you might hear, you know a group (. ) say you’ve got a singer with a band, you might hear the band playing from one place and the voice coming from another (. ) like, you are not going to go through every individual instrument in the band and say - oh well they are all coming from here
10: No, ah but we are only talking about three, two instruments and a voice aren’t we?
17: Yes
R: At the moment, yes
17: I mean if somebody hears them as three individual, then, put them down as three (. ) if they hear it is one put it down as one, everybody is different, we all hear things in a different way.
R: So we could have a possible, the only thing I am really checking is, could we have this possible thing occurring where you have this instrument ((drawing)) or set of instruments there ((on one side of the vehicle)), this instrument here, which is next to this instrument here ((on the other side of the vehicle)), Is that ever going to be the case where you would put those two ((sets of instruments)) together?
8: Well that’s what, yes, that’s what you were describing ((to ‘17’)) isn’t it, if that does happen?
17: Yes, I mean I’d say, if you heard that, if those two ((instruments together)) seemed to be coming from the same source and you do that as one only (. ) with width and depth. And that one ((the single remaining instrument)) remains a single source, because you hear them differently. But you hear that ((the two instruments together)) as a combined sound.
Excerpt 4 Panel III (Meeting 2) - discussion about height

3: Height?
R: Height?
3: Hmmm
R: How about height?
3: Hmmm, height (.) that was something I mentioned yesterday. Whether it was coming sort of, from above (.) from above the dashboard =
17: = but wouldn't that just be a function of where you are able to stick the speakers in the vehicle? (.) I mean if you could stick a door woofer - instead of being down at the bottom as traditionally they are - up at the top and then they ((the speakers)) would all be at the same level, they'd all be like ear level or eye level. But physically it is impossible to put it in the car. So does height (.) affect much?
3: I don't know, I found myself (.) listening to the foot-well (.) on some of those
((general - good humoured - laughter from remaining panelists))
3: My, my attention was specifically towards the passenger foot-well
10: So you'd like your speaker in the floor (apparently)!
3: Well, yes!
8: God that foot-well sounds good! ((more laughter))
3: But, but the sound was coming from there (.) as opposed to coming sort of (.) =
17: = yes =
3: = above the dash, I could, I could distinguish that quite easily
17: But you can't put a - well you could put - you generally put a centre a fill at eye-level (.) and the door speaker, kind of foot-well level wouldn't you?
3: mmm
((Brief discussion about speaker location within vehicle and ramifications for reproduced audio))
17: . hhhh I, I don't think I would like to include height (.) myself
R: Do we want to leave it in there until we've had a listen? ((panel is about to trial their language, describing the current source material reproduced inside the vehicle))
8: Yes, I think so
3: Height is also an el..., its like also an element of the surround isn't it?
17: Could be
((pause in discussion))
8: How would you describe height then?
3: Its just the z axis isn't it? =
17: = big and little? =
3: = in terms of the surround
10: High, low?
3: (oh right, ok, yes) high or low
8: Its very hard isn't it? Cause it is there, or is it there ((referring to a graphical response sheet))?
3: I found it very easy, like if my head like went to the floor ( ) basically
17: I noticed sometimes that you could hear the sound coming from there (door speaker) but I put that down to the fact that that was where the speaker was located, and that is how it is going to be
3: But isn't that the same as saying its specifically on the (barrier) but it is coming from over there because it is the speaker over there, its the same thing isn't it?
17: I don't know, if you go to the cinema and something crashes over your head, do you generally go ((demonstrates looking up)) (if the speaker's above you, or do you hear =
3: = everything's above me!
8: Yes we will leave it ((height)) in, but what words can we use for it (.) high and low?
3: High and low
Excerpt 5 Panel II (Meeting 2) - discussion about height

14: The other thing is, is the height we mentioned (.) and how you'd depict that. The only thing I can think of is maybe have a figure - say 'one' to 'three' - so 'one' stays at the top, 'two' is in the middle and 'three's on the bottom
9: (.) I think that's (.) over complicated =
2: =hbb, uh I'd prefer not to bother with height
14: I can't see you can (.) How can you ignore it though? I can't. I couldn't ignore it! Perhaps I am overcomplicating (.)
R: Could you have a word (scale) for height?
14: You could, you could (.) [it could be 'A', 'B', 'C' or whatever it is
9: maybe you need two pictures? (.) Maybe you need two, maybe you need a top view and a side view?
2: Well, or just a comment if you want to put a comment on (.) [you know?
14: But you know like you've got speakers up in the top post or something, you may just hear that (.) Well obviously it's going to come (.) there is a difference in the height isn't there? (.) Or you may have, you may have a speaker on the floor, that you just hear a sound from
2: Yes, but what I am saying is you can, a single instrument can have different heights [depending on
9: yes, as you go through the [frequency bands
2: = the frequency that it is operating at (.) th
14: What, so it may go across ( )
9: door to door =
2: = yes well, I can't see, you'd have to have a three dimensional[ ( )
14: oh alright, ok
R: Could you do it with two, [two-dimensional?]
2: I can't see how you could
9: What two, two-dimensional pictures?
2: Yes
9: You [have it like that ((plan)) and then you have it that way? ((elevation))
14: Yes I think you could
2: How?
9: Well basically ((
2: you show me a picture
14: well by, by putting a [by putting a figure on it like we said: 'One', 'Two', 'Three' ((drawing))
9: basically, what you'd do is, in this picture you'd depict the area looking from above where it is, and then from that side you'd depict how high in the car it is
2: No, I can't (.)
14: If you heard it (.) from the floor to the middle and the top - right the way through - then you'd put the figures 'one', 'two' and 'three' down. If you just heard it from the bottom (.) perhaps you'd just put the 'one' in, you could put, you know, you could have a combination =
((2 talks for a while about why he doesn't think 'height' can be depicted graphically...))
2: But you can't depict it ((height))!
9: But that's what I am saying, if you had it [ on this
14: but you can't =
2: = if you had another [drawing you could
9: = if you had on this, another drawing (.) of the car, blah, blah, blah ((drawing)) you could then put (.) right, you've put the position (.) and then you could put the height in the car (.) And because its like another section of the car =
2: = so, so what if we had three then? and, and a centre?, where are we going to put height on those then?
9: Well I still think you can do that
2: How? (.) Because it will all be in the same plane?
9: Different colours?((general laughter))
R: Um, we are agreed that height is important though [yes?
14: I think so yes, I don't think you can ignore it
Excerpt 6 Panel IV (Meeting 1) - discussion about area

R: Can you explain your picture to other people?
13: That's where the instrument, I found, well thought where the instrument was (central circle) and that's the, how far the sound, the way the sound was coming out, projecting out (lines around the central circle)
R: Has anyone else dealt with projection?
((Pause))
7: Yes
16: No
7: ((Looking through own graphical responses)) That was (.) often they were circular but (.) as well as the width (.) then (.) basically the boundary is the overall (.) so it is projecting (.) That's more of an example: ((shows his circular description to remainder of panel)) So it is projecting (.) that way ((front-to-back)) (.) It has some depth as well ( )
R: Depth? So would you say, depth is the thing that goes from front-to-back in the instrument yes?
7: = (mmmm) =
R: = would you say that, that ((looking at the outer part of '13's' graphical descriptor)) (.) could be a similar thing?
13: Yes (.) yes
R: ((to '19')) Are you dealing with depth at all?
19: Uh (.) not (.) well (.) in a, in a way yes, but that's, that's what I'd say ((indicating on graphical response)) there was my depth (.) Depth towards, yes, distance from the focal point to where I perceived the boundary to be, as it were.
R: Right, ok. So you are all pretty much, with the exception of '16' who has got very simple 'width' and 'direction'
16: Cause you ask me, how deep is a cassette, uh sound say, and I'd say: Does it really sound like its got any depth?
((pause))
R: Ok, '19' carry on down your list ((individual verbal terminology list)) so we can just (get an idea ( )
19: Uh, the area of the sound, which is basically what is encompassed in there ((inside the triangle)) (.) so that's where '7' has got a pocket on his (.) and '13' has got like a =
13: = area
19: Area on hers as well
R: So you are happy with using 'area' as a descriptive term for what you are up to?
13: Yes
16: Mmm, to a point
R: Yes? explain your reservations
16: Well in my particular case, people say they hear depth to it, sorry I don't really
R: You hear it as a flat sound?
16: Yes
R: Right
7: I use the word envelope and that's got (.) basically two dimensions
19: Yes, I use that later on actually, well there are another two here 'R' which are nigh on the same: I've got pocket of sound which is the same as area I suppose, as is envelope of sound
Excerpt 7 Panel I (Meeting 2) - discussion about ambience

11: What did we describe ambience? We called it um (.) scattering, reverberation

((pause))

4: I dunno, maybe a bigger shape in two different colours. The ambience being (.) darker or lighter, or something

R: Describe ambience then, as a group, come up with a definition of this word ambience so that I know you are all using it the same (.) A description that you can come to a consensus on.

11: Ambience to me is um (.) live sound (.) um (.) a sound that fills the space, um (.) reverberation, um (.) a feel

R: What's its opposite, or what are the opposites on a scale of ambience?

11: Dry sound, um (.) dead (.) devoid of life

((general muttering))

R: What's on the other side of this? So if you have dead and dry

11: If you have dead and dry, if you want to describe a dead and dry sound (.) um

20: Its smaller, more located to the (.) this point source (.) that's how I see it (.) At home, if you have a stereo recording and you have, you can hear all the instruments but just from the speakers (.) that's quite dead, but with ambience and stuff you (.) you get a picture of-

11: = of placement

20: Yes, a little bit more (around it)

4: (.) Surely we are measuring that (.) with length and the breadth?

20: No no, no, the length and width of the (.) instrument is not the same (.) as far as I see it anyway

4: I understand what you are saying ( )

11: it is all to do with the space it was recorded in and where it is played (back

20: yes, exactly =

11: = like you could have a (.) where it was recorded could be in a, in a very dead space (.) But you could still play it back and still have a (.) a relatively (.) live or ambient feel (.) depending upon where (.) if you had a number (.) a number of locations where it was coming from. (.) There would still, there might be some dry characteristic to the sound but, you might get a feeling for (.) a space =

20: = oh yes (.) if you were playing it (.) in a warehouse with a load reverberation [and stuff

11: Yes, yes, and then you've got the opposite

4: So how do you draw that then, in this car?!

20: The warehouse?

((general nervous laughter))

5: Shade in the area?

((R provides the panelists with more blank car representations and more pens for drawing))

11: Scattering? (.) The scattering of sound maybe? A live =

20: = I am not very happy with (.) the word scattering

11: Well I mean, its, its all other words that you'd use to describe ambience (.) But you might say that ambience is the (.) the word that we all agree on. But there might be other meanings. Just to help describe ambience really

... (("11" is busy drawing "ambience")

4: ( ) That's good!

11: Ambience!

R: So something not straight (.) not, not (.) defined [as such]? ((referring to the drawing '11' is doing))

11: Yes (.) you know just something that um (.) well we've got to think of something haven't we!

((Everyone laughs))

R: And would that be a whole (.) shape, or would you have a lot of those squiggly (.) lines?

11: It could be any shape and it could be any size and (.) in any location (.) but (.) well, maybe not location but any shape and size =
Clarifying individual listeners' descriptions and developing U-GAL

Chapter 3 - Appendices

4: = So, all we've done really then is this, there is '20's' shape isn't it? ((draws a smooth line over the top of '11's' squiggly line)) Like that? All we've done is just taken (.) the edges and made them squiggly, effectively!

11: Yes and you could describe width in that, and depth

4: = what would be wrong with, lets say, having a shape ((draws a shape using multiple straight lines to define it's dimensions)) like this (.) and then having the ambience outside that? Lets say that's like that ((draws some squiggly lines)) and then maybe using a highlighter to go on outside it ((uses highlighter to outline shape))?

11: Yes

5: So are you saying that (.) I've drawn this (.) large circle now (.) are you saying I've heard a lot of ambience (.) in this and I can't lo.. (.) you know, I can't say its coming from here, here, or here, I've just drawn it in. Whereas (.) if I've drawn something really small, then I am,  

11: Possibly, yes, yes

5: So anyway, instead of drawing something huge like that, I'd say it was a bit by here, bit by here or bit by here, ((pointing to distinct positions within an ensemble)) rather than saying, well I don't know ( )

R: But an ambience in there as opposed to?

5: As opposed to being an area of confusion, then I'd say it was looking like there was something here, here ( )

R: So would you have a very (.) very narrow width and a very narrow (.) depth and then just (.) and then just a blob of ambience?

5: Must have got confused see!

20: Ha ha!

11: You are describing two things there,

5: Yes, yes I can see now

11: You know, you see where your 'C' ((letter representing the centre of the 'cello' instrument)) is (.) you know you could say there is width and depth and location there (.) but (.) you may have also drawn, maybe what you have also interpreted as ambience? =

5: = ambience, yes

R: It's a good get out clause!

20: Yes, the blob of confusion! It's a, a big black, like a circle that you put over there (.) a get out of jail card

((general laughter))

R: ((Bringing proceedings back to order)) But we have to know (.) is ambience distinct from this confusion? Is ambience distinct from confusion, is ambience a certain thing that is related to the liveness or the deadness of a sound?

20: Ambience at least for me could be a little bit of (both)

((Tape runs out & re-starts))

11: It is almost between live and dead sound (.) You know? (.) Is it (.) can you see (.) the, the sound like a live band playing (.) or (.) does it seem (.) um, studio and very precise (.) a studio sound or a very precise sound (.) with no feel? [with feel, without feel?]

20: For a single instrument (.) it is hardly ever (.) confusion, I think, its more ambience. But when you start to mix it all

11: But you can also get ambience (.) for feel of space, depending upon (.) the number of locations you've got in the car (.) and you've got a feel for a space that is maybe bigger than the environment you listening to
Excerpt 8 Panel I (Meeting 3) - discussion about *ambience*

11: So, what is another word for ambience then? (.) say, um =
4: = I think live sound is good, I do
11: Yes, live sound, yes (.) we've used that
4: Close to the point, yes (.) I'd like to stick with that, I think that's, that's pretty good
11: I agree
((pause))

R: So we will use that as a (.) could there be any ambiguities in the use of live sound? What kind of description of live sound would you use?

11: I think live sound might be the extreme (.) where you say that, that's ambience (.) um, an opposite to live sound would be (.) um, a dead sound, or a um (.) synthetic sound or something like that

R: Where you've used live sound, where is it? (.) (looking back through transcript of previous meeting)

20: Its on the second page in the middle

R: Live sound ((reading)) 'feeling that you are involved in the sound', do you think that that is a useful definition (.) or is that a different attribute altogether?

11: I think what you (.) what (.) ambience is, is (.) its a feeling of (.) you are in a live gig, say, a live music concert, whatever type of music it is, you get that feeling of, of (.) um, music and *sound* (.) that you are experiencing live music (.) Whereas if it was, if it was (.) the more reproduced it is, the more (.) um, or the (.) how the space (.) sounds to you, if its, if its less like (.) uh, sounds coming from all around you and it is coming from one direction then its (.) its got less of a feeling of, of a concert and that's how I, [I

4: I agree with you entirely '4', in the vehicle that is what they want to reproduce isn't it?

20: Yes

R: So we want the feeling of experiencing live music then?

20: Uhhu
11: Yes

R: (.) Anything else you want to add to live sound to make it, make the point?

11: Um (.) quality? (.) um (.) *naturalness* isn't it! It's a natural sound

4: Well, what happens if the listener hasn't *experienced* a live stage? They are not going to be able to (define the live stage) =

11: = well he has heard something, whenever he walks into a building, (he hears
5: it is still all around you though isn't it?

20: = well, you can always

11: if you walk into a building you are going to (hear people talking

20: you are live, just hearing you taking, if, if you sit in the car and get a feeling of a person really sitting on the dashboard talking to you, that's (.) a little bit of a 'live' sound

R: Take on board '4's points that, if you haven't been to an orchestra, or, you don't know what it sounds so

11: But its that feeling of space isn't it? so (.) you know, you could walk into one room and its very dry and dead (.) but the next room you could have a very *live* and reverberant room

20: But it could, it could go beyond natural reverberation, so where are we then?

11: Beyond natural reverberation?

20: Yes, for a concert hall for example, if the concert hall were to be a reverberant space, then its not going to be an ambient live sound

4: So how do you measure that? (.) You know, the biggest concern I've got is how do you measure ambience? Everything else we can more or less cope with, but, uh, trying to tackle this one is going to be the biggest problem I think.

20: Yes to try and put that 'live' feeling into an arrow!
Excerpt 9 Panel II (Meeting 1) - discussion about *ambience*

2: I think we are agreed that ambience is *something*

9: Yes

14: Ambience is one end of a scale to me

6: Ambience is everywhere

2: We need to find out how ambient it is

9: It's not the end of a scale, I disagree with you [*14’]

6: *Around?*

9: I think there should be a separate scale for ambience, it's either ambient or it's not ambient!

6: Dead? Not ambient is a dead sound, not moving?

2: Well does it have ambience or not? Let's just boil it down to that?

14: Mmm ah no, what is ambience?

9: Ambience is like the [feeling of space]

6: *Around everywhere*

9: Do you feel like you are sitting in this little car, or do you feel like you are sitting in a hall?

14: No that is not ambience to me, ambience is another sound or like a background sound, someone coughing over there, or some shuffling =

2: = Ah that is not ambience

9: Its part of ambience

2: No, What he's saying is if someone coughs behind [you]

14: Ah well its difficult then, that's a specific point, I know what you are saying

2: That's a point source behind you =

9: = But it adds to the ambience

R: When you said ‘the feeling of space’

14: That is what it is

R: Can someone coughing behind you come into a feeling of space?

14: Of course, that is what generates a feeling of space

9: That is what I mean, it is part of ambience

14: Say you've got 2000 people coughing all around you

...  

14: Are you saying that ambience cannot be positional, A cough or something?

9: That adds to ambience, but it is not ambience

14: To me, we are in this room and I can hear that thing and I think that's ambience

9: No it adds to ambience, the reflections off the wall behind you, you don't get much [ambience

14: No I don't say you get much but you do get something in the background definitely

9: You get a lot more ambience in a church hall or something, like the echo and stuff (;) in a big space
Excerpt 10 Panel II (Meeting 2) - discussion about *ambience*

14: I've got a problem with ambience because I don't understand it
9: Because it's the end of a scale isn't it?
14: My interpretation of ambience is different from everybody else. To me, ambience is
R: It's not different to everybody else, it's different from =
14: = It's different from these two ((referring to '2' and '9'))
2: From what, from?

2: So, yes but, you, you know () find a () what other () ambience is a term that we use, I know where you are coming
from, you () you are taking the literal meaning of ambience to mean, something like a mood even, like if you go into
a club and say that's got ambience, yes?
14: Well () hhhh [yes
2: I think
14: Well no, I am talking about sound () I'm not talking about, you know, anything else. I am purely talking about sound,
but to me its just background, its, its () background [sound
R: Ok, you (('14')) would say that you could pick up one of these ((existing shapes)) shove it at the back of the car? Or,
would you have to describe an intensity level, something to say that it is background?
14: No, uh, I, I, there is only () To me there is only one () there would be one symbol for ambience () for
2: So you think its, that all ambience is the same size?
14: Yes
2: And its full?
14: It covers yes, it would cover the space () but you see, if its not, if its positional or directional () to me then it
wouldn't be ambience
2: Yes that's true
CHAPTER 4 APPENDICES
Appendix 4.A.1 Extracts from the U-GAL UI tutorial

Figure 4.A.1 Tutorial 1st page. Introduction to the screen
Thank you for taking the time to help us today.
This short guide will introduce you to the screen you are holding and show you how to use it.
This screen is a "touchscreen". You can press buttons on the screen just by touching them with your pointer.
Try this now: Press and release the "take and control" button to continue.

Figure 4.A.2 Tutorial 2nd page. Introducing descriptors
You can move shapes around on this screen. Do this by holding the pointer down on a shape and dragging it across the screen.
Have a go at moving around the shapes on the right. Familiarise yourself with moving them back and forth, up and down. Also, try moving the C, V and P. Once you think you’ve got the hang of it, press "take and control" to continue.

Figure 4.A.3 Tutorial 3rd page. Moving descriptors
This time, drag the outlines from the right onto the red template on the left. Try to place the shapes exactly on top of the template. You will be able to continue once you have correctly placed all the outlines over their red counterparts. If you get in a mess, you can re-start this screen using the "take and control" button in the top right corner.

Figure 4.A.4 Tutorial 5th page. Changing descriptor size
Now try matching the two shapes on the right to the red template on the left. You will need to change their shape, and move them into place. You will be able to continue once you have correctly placed both shapes.

Figure 4.A.5 Tutorial 7th page. Creating descriptors
From now on, you will have to create your own shapes. To create shapes, just tap the shape you need in the box at the top of the screen. Try this now: See if you can create one of each shape type. Make a, b, c, d, e, f, g, h, i and j.
Once you’ve created the shapes, you can move them around and trash them just like before.
Press "take and control" when you are ready to move on.

Figure 4.A.6 Tutorial last page. Combining all actions
You’ve almost finished! This time, using the skills you’ve learnt so far, you need to: 1) Create the shapes you need to match the red template below.
2) Adjust the shape sizes accordingly.
3) Move the shapes so that they fit over the template.

Thank you for taking the time to help us today.
Figure 4.A.2 Tutorial 2nd page. Introducing descriptors
You can move shapes around on this screen. Do this by holding the pointer down on a shape and dragging it across the screen.
Have a go at moving around the shapes on the right. Familiarise yourself with moving them back and forth, up and down. Also, try moving the C, V and P. Once you think you’ve got the hang of it, press "take and control" to continue.

Throughout this tutorial you will often see useful tips down here. (Tip: You will find this easier if you press quite hard with the pointer. Don’t worry about damaging the sensor.)
Appendix 4.A.2 Instructions for listeners using U-GAL

Using the provided shapes (□ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ ^}{EX}
Appendix 4.B Individual listener consistency plots

Appendix 4.B overview

Graphical plots within this appendix indicate the level of consistency achieved by individual listeners when using U-GAL descriptors to describe their auditory spatial experiences. Plots display all U-GAL descriptors used by a listener when representing a single stimulus. Each figure was constructed by re-plotting together all responses provided by a listener to three repeated runs of the same stimulus. Eight plots are included for each listener. Visual plots for two listeners are presented on the same page, each page containing plots numbered from 1 - 16. Each page is laid out as follows:

Plot 1 and 9: **Percussion left**
- Percussion amplitude panned far left in the two-channel stereo scene

Plot 2 and 10: **Percussion central**
- Percussion amplitude panned central

Plot 3 and 11: **Cello left**
- Cello amplitude panned far left

Plot 4 and 12: **Cello central**
- Cello amplitude panned central

Plot 5 and 13: **CPV Narrow**
- Cello, Percussion, Voice (closely spaced around centre of the stereo scene)

Plots 6 and 14: **VCP Wide**
- Voice (left), Cello (central), Percussion (right)

Plots 7 and 15: **VCP Left**
- Voice (left), Cello (centre left), Percussion (central)

Plots 8 and 16: **CPV Right**
- Cello (central), Percussion (centre right), Voice (right)

An instrument’s position (left to right) within an ensemble is the same as the ensemble’s name. E.G. the ensemble **CPV Narrow** positions the cello left of the percussion and the voice right of the percussion, with all sources closely spaced around a central location in the two-channel stereo scene.

When pages from this appendix are referred to in the main text they will be preceded by the figures 4.B. Hence page 1 from this appendix will be referred to as 4.B.1.
4. B. 1 Agreed use of U-GAL descriptors

LISTENER 2 (experienced existing listener and participant in U-GAL final development meetings)

1. Percussion Left
2. Percussion Central
3. Cello Left
4. Cello Central
5. CPV Narrow
6. VCP Wide
7. VCP Left
8. CPV Right
9. Percussion Left
10. Percussion Central
11. Cello Left
12. Cello Central
13. CPV Narrow
14. VCP Wide
15. VCP Left
16. CPV Right
4.B.2 Agreed use of U-GAL descriptors

LISTENER 1 (experienced new listener)

1. Percussion Left
2. Percussion Central
3. Cello Left
4. Cello Central
5. CPV Narrow
6. VCP Wide
7. VCP Left
8. CPV Right

LISTENER 2 (inexperienced new listener)

9. Percussion Left
10. Percussion Central
11. Cello Left
12. Cello Central
13. CPV Narrow
14. VCP Wide
15. VCP Left
16. CPV Right
4.B.3 Some individuality in U-GAL descriptor use

LISTENER 3 (inexperienced existing listener)

1. Percussion Left
2. Percussion Central
3. Cello Left
4. Cello Central
5. CPV Narrow
6. VCP Wide
7. CPV Left
8. CPV Right
9. PV Narrow
10. PV Wide
11. Cello Left
12. Cello Central
13. CPV Narrow
14. VCP Wide
15. CPV Left
16. CPV Right
4.B.4: Some individuality in U-GAL descriptor use

LISTENER 6 (inexperienced new listener)


LISTENER 21 (inexperienced new listener)

4.B.5 Individuality in U-GAL use

LISTENER 18 (experienced existing listener)

1 Percussion Left 2 Percussion Central 3 Cello Left 4 Cello Central 5 CPV Narrow 6 VCP Wide 7 VCP Left 8 CPV Right

LISTENER 9 (experienced existing listener)

9 Percussion Left 10 Percussion Central 11 Cello Left 12 Cello Central 13 CPV Narrow 14 VCP Wide 15 VCP Left 16 CPV Right
4.B.6 Individuality in U-GAL use

LISTENER 11 (experienced existing listener and participant in U-GAL final development meetings)

1 Percussion Left  2 Percussion Central  3 Cello Left  4 Cello Central  5 CPV Narrow  6 VCP Wide  7 VCP Left  8 CPV Right

LISTENER 15 (inexperienced existing listener and participant in U-GAL final development meetings)

9 Percussion Left  10 Percussion Central  11 Cello Left  12 Cello Central  13 CPV Narrow  14 VCP Wide  15 VCP Left  16 CPV Right
4.B.7 Individuality in U-GAL use

LISTENER 7 (musician and new listener)

1 Percussion Left  2 Percussion Central  3 Cello Left  4 Cello Central  5 CPV Narrow  6 VCP Wide  7 VCP Left  8 CPV Right

LISTENER 16 (inexperienced new listener)

9 Percussion Left 10 Percussion Central 11 Cello Left 12 Cello Central 13 CPV Narrow 14 VCP Wide 15 VCP Left 16 CPV Right
4.B.8 Individuality in U-GAL use

LISTENER 20 (experienced new listener)

1. Percussion Left
2. Percussion Central
3. Cello Left
4. Cello Central
5. CPV Narrow
6. VCP Wide
7. VCP Left
8. CPV Right

LISTENER 8 (musician and new listener)

9. Percussion Left
10. Percussion Central
11. Cello Left
12. Cello Central
13. CPV Narrow
14. VCP Wide
15. VCP Left
16. CPV Right
4.B.9 Further examples of visual plots for individual listeners

**LISTENER 5 (experienced new listener)**


**LISTENER 17 (inexperienced new listener)**

4.B.10 Further examples of visual plots for individual listeners

LISTENER 19 (inexperienced existing listener)
Appendix 4.C Overlaid descriptor plots

Appendix 4.C overview

Appendix 4.C contains graphical plots illustrating how the individual U-GAL descriptors were used to represent the spatial characteristics of different items of programme material. In contrast to the individual listener plots of appendix 4.B, this appendix plots the descriptions of all listeners together where the same U-GAL descriptor appears to have been used to represent a similar experience. U-GAL descriptors were developed to represent a listener’s experience as outlined in Appendix 4.A

An instrument’s position (left to right) within an ensemble is the same as the ensemble’s name. E.g. the ensemble CPV Narrow positions the cello left of the percussion and the voice right of the percussion, although all sources are closely spaced around a central location in the two-channel stereo scene.

When pages from this appendix are referred to in the main text they will be preceded by the figures 4.C. Hence page 1 from this appendix will be referred to as 4.C.1.
4.C.1 U-GAL letter plots

Plots 1-8: Use of letters within FOCAL descriptors

- 1: Percussion Left
- 2: Percussion Central
- 3: Cello Left
- 4: Cello Central
- 5: CPV Narrow
- 6: CPV Wide
- 7: VCP Left
- 8: CPV Right

Plots 9, 16: Use of letters within AREA descriptor

- 9: Percussion Left
- 10: Percussion Central
- 11: Cello Left
- 12: Cello Central
- 13: CPV Narrow
- 14: CPV Wide
- 15: VCP Left
- 16: CPV Right
4.C.2 FOCAL plots (solo instruments only)

Plots 1-4 Use of FOCAL without AREA

1 Percussion Left

2 Percussion Central

3 Cello Left

4 Cello Central

Plots 5-8 Use of FOCAL as a localisable centre/core within AREA

5 Percussion Left

6 Percussion Central

7 Cello Left

8 Cello Central
4.C.3 AREA and FOS plots (solo instruments only)

Plots 1-4 Use of AREA

Plots 5-8 Use of FOS without AREA

1. Percussion Left
2. Percussion Central
3. Cello Left
4. Cello Central
5. Percussion Left
6. Percussion Central
7. Cello Left
8. Cello Central
4.4.4 AREA Plots (ensembles by instrument)

Plots 1-3: Cello Percussion Voice narrow ensemble

Plots 4-6: Voice Cello Percussion wide ensemble

Plots 7-9: Cello Percussion Voice right ensemble

Plots 10-12: Voice Cello Percussion left ensemble

1 Cello (at centre)
2 Percussion (at centre)
3 Voice (at centre)
4 Voice (left)
5 Cello (at centre)
6 Percussion (right)
7 Cello (at Centre)
8 Percussion (centre-right)
9 Voice (right)
10 Voice (left)
11 Cello (centre-left)
12 Percussion (at centre)
4.5 C. Plots of ensembles by instrument where FOS was used without AREA

Plots 1 – 3 Cello Percussion Voice narrow ensemble

Plots 4 – 6 Voice Cello Percussion wide ensemble

1 Cello (at centre)
2 Percussion (at centre)
3 Voice (at centre)

4 Voice (left)
5 Cello (at centre)
6 Percussion (right)

Plots 7 – 9 Cello Percussion Voice right ensemble

Plots 10 – 12 Voice Cello Percussion left ensemble

7 Cello (at Centre)
8 Percussion (centre-right)
9 Voice (right)

10 Voice (left)
11 Cello (centre-left)
12 Percussion (at centre)
4.C.6 Plots of ensembles by instrument where FOCAL was used within an AREA

Plots 1 – 3 Cello Percussion Voice narrow ensemble

Plots 4 – 6 Voice Cello Percussion wide ensemble

1 Cello (at centre)  2 Percussion (at centre)  3 Voice (at centre)  4 Voice (left)  5 Cello (at centre)  6 Percussion (right)

Plots 7 – 9 Cello Percussion Voice right ensemble

Plots 10 – 12 Voice Cello Percussion left ensemble

7 Cello (at Centre)  8 Percussion (centre-right)  9 Voice (right)  10 Voice (left)  11 Cello (centre-left)  12 Percussion (at centre)
### Appendix 4.D Statistical graphs and tables

Table 4.D.1 Extracts from the Analysis of Variance table for the variable stimuli, including an estimate of effect size

<table>
<thead>
<tr>
<th>Listener</th>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta²</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>STIMULI</td>
<td>8930.000</td>
<td>3</td>
<td>2976.667</td>
<td>26.918</td>
<td>.000</td>
<td>.910</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>884.667</td>
<td>8</td>
<td>110.583</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>9814.667</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>STIMULI</td>
<td>10938.250</td>
<td>3</td>
<td>3646.083</td>
<td>35.717</td>
<td>.000</td>
<td>.931</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>816.667</td>
<td>8</td>
<td>102.083</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>11754.917</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>STIMULI</td>
<td>15867.000</td>
<td>3</td>
<td>5280.000</td>
<td>16.857</td>
<td>.001</td>
<td>.863</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>2510.000</td>
<td>8</td>
<td>313.750</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>18377.000</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>STIMULI</td>
<td>23602.917</td>
<td>3</td>
<td>7867.639</td>
<td>158.942</td>
<td>.000</td>
<td>.983</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>396.000</td>
<td>8</td>
<td>49.500</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>23998.917</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>STIMULI</td>
<td>16805.667</td>
<td>3</td>
<td>5601.889</td>
<td>9.781</td>
<td>.005</td>
<td>.786</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>4582.000</td>
<td>8</td>
<td>572.750</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>21387.667</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>STIMULI</td>
<td>22668.667</td>
<td>3</td>
<td>7556.222</td>
<td>8.004</td>
<td>.009</td>
<td>.750</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>7552.000</td>
<td>8</td>
<td>944.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>30220.667</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>STIMULI</td>
<td>30947.729</td>
<td>3</td>
<td>10315.910</td>
<td>536.472</td>
<td>.000</td>
<td>.955</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>153.833</td>
<td>8</td>
<td>19.229</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>31101.563</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8a</td>
<td>STIMULI</td>
<td>3501.283</td>
<td>3</td>
<td>1167.094</td>
<td>1.997</td>
<td>.193</td>
<td>.428</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>4675.720</td>
<td>8</td>
<td>584.465</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>8177.003</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8b</td>
<td>STIMULI</td>
<td>11592.401</td>
<td>3</td>
<td>3834.134</td>
<td>3.224</td>
<td>.082</td>
<td>.547</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>9514.002</td>
<td>8</td>
<td>1193.250</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>21016.402</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>STIMULI</td>
<td>28396.667</td>
<td>3</td>
<td>9465.556</td>
<td>46.801</td>
<td>.000</td>
<td>.946</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>1618.000</td>
<td>8</td>
<td>202.250</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>30014.667</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>STIMULI</td>
<td>8513.667</td>
<td>3</td>
<td>2837.889</td>
<td>12.488</td>
<td>.002</td>
<td>.824</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>1818.000</td>
<td>8</td>
<td>227.250</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>10331.667</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>STIMULI</td>
<td>30446.000</td>
<td>3</td>
<td>10148.667</td>
<td>234.200</td>
<td>.000</td>
<td>.989</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>346.667</td>
<td>8</td>
<td>43.333</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>30792.667</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>STIMULI</td>
<td>12316.667</td>
<td>3</td>
<td>4105.556</td>
<td>58.704</td>
<td>.000</td>
<td>.957</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>614.000</td>
<td>8</td>
<td>76.750</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>14130.667</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>STIMULI</td>
<td>16496.250</td>
<td>3</td>
<td>5498.750</td>
<td>94.806</td>
<td>.000</td>
<td>.973</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>464.000</td>
<td>8</td>
<td>58.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>16960.250</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>STIMULI</td>
<td>31875.583</td>
<td>3</td>
<td>10625.194</td>
<td>196.460</td>
<td>.000</td>
<td>.987</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>432.667</td>
<td>8</td>
<td>54.083</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>32308.250</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>STIMULI</td>
<td>9378.917</td>
<td>3</td>
<td>3126.306</td>
<td>34.074</td>
<td>.000</td>
<td>.927</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>734.000</td>
<td>8</td>
<td>91.750</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>10112.917</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>STIMULI</td>
<td>104.917</td>
<td>3</td>
<td>34.972</td>
<td>397</td>
<td>.759</td>
<td>.129</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>705.333</td>
<td>8</td>
<td>88.167</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>810.250</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>STIMULI</td>
<td>18433.583</td>
<td>3</td>
<td>6144.528</td>
<td>15.953</td>
<td>.001</td>
<td>.857</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>3081.333</td>
<td>8</td>
<td>385.167</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>21514.917</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>STIMULI</td>
<td>28167.333</td>
<td>3</td>
<td>9389.111</td>
<td>8.643</td>
<td>.007</td>
<td>.764</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>8690.667</td>
<td>8</td>
<td>1086.333</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>36858.000</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>STIMULI</td>
<td>2605.583</td>
<td>3</td>
<td>868.528</td>
<td>2.780</td>
<td>.110</td>
<td>.510</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>2499.333</td>
<td>8</td>
<td>312.417</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>5104.917</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>STIMULI</td>
<td>73044.250</td>
<td>3</td>
<td>24348.083</td>
<td>29.582</td>
<td>.000</td>
<td>.917</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>6584.667</td>
<td>8</td>
<td>823.083</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>79628.917</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>STIMULI</td>
<td>26580.667</td>
<td>3</td>
<td>8860.222</td>
<td>133.404</td>
<td>.000</td>
<td>.980</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>531.333</td>
<td>8</td>
<td>66.417</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 4.D.1 Graph illustrating the proportion of the total variability in described location that can be accounted for by the variation in the independent variable stimuli. Reference line denotes 50% of the total variability.

Table 4.D.2 Table of non-significant differences from post hoc comparisons of programme material location. (* indicates results were obtained using the Games-Howell and not the Bonferroni procedure)

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Number of listeners</th>
<th>Listener ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percussion Left and Percussion Central</td>
<td>4</td>
<td>8a (p = .422), 16 (p = .842)<em>, 19 (p = .493), 22 (p = .052)</em></td>
</tr>
<tr>
<td>Percussion Left and Cello Central</td>
<td>10</td>
<td>1 (p = .347), 5 (p = .684), 6 (p = .242), 10 (p = .234), 13 (p = 1.000), 16 (p = .859)<em>, 17 (p = .390)</em>, 18 (p = .416)*, 19 (p = .224)</td>
</tr>
<tr>
<td>Cello Left and Percussion Central</td>
<td>5</td>
<td>8a (p = .383), 15 (p = .200)<em>, 16 (p = .865)</em>, 19 (p = 1.000), 22 (p = .723)*</td>
</tr>
<tr>
<td>Cello Left and Cello Central</td>
<td>13</td>
<td>1 (p = 1.000), 5 (p = .057), 6 (p = .330), 9 (p = .062)<em>, 8a (p = 1.000), 10 (p = .878)</em>, 11 (p = .130), 15 (p = .234), 15 (p = 1.99)<em>, 16 (p = .994)</em>, 17 (p = .257)<em>, 18 (p = .565)</em>, 19 (p = .511), 22 (p = .675)*</td>
</tr>
</tbody>
</table>
The following graphs (figure 4.D.2 - 4.D.15) illustrate means and 95% confidence intervals for individual listeners when describing the various positions of the different stimuli. In the graphs, the centre of the vehicle (positioned at 0) is indicated by the reference line, the left boundary of the vehicle is referred to by and the right boundary of the vehicle is denoted by . Graphs are displayed with units of pixels.

Figure 4.D.2 Means and 95% CI for listener 13 when describing the position of different stimuli

Figure 4.D.3 Means and 95% CI for listener 1 when describing the position of different stimuli

Figure 4.D.4 Means and 95% CI for listener 22 when describing the position of different stimuli

Figure 4.D.5 Means and 95% CI for listener 15 when describing the position of different stimuli

Figure 4.D.6 Means and 95% CI for listener 17 when describing the position of different stimuli

Figure 4.D.7 Means and 95% CI for listener 18 when describing the position of different stimuli
Figure 4.D.8 Means and 95% CI for listener 9 when describing the position of different stimuli

Figure 4.D.9 Means and 95% CI for listener 6 when describing the position of different stimuli

Figure 4.D.10 Means and 95% CI for listener 5 when describing the position of different stimuli

Figure 4.D.11 Means and 95% CI for listener 10 when describing the position of different stimuli

Figure 4.D.12 Means and 95% CI for listener 16 when describing the position of different stimuli

Figure 4.D.13 Means and 95% CI for listener 19 when describing the position of different stimuli
Figures 4.D.14 and 4.D.15 compare responses for different listening groups when solo instrument source location is described. In both means and 95% confidence interval graphs, the centre of the vehicle (0) is represented by the reference line............

Figures 4.D.16 and 4.D.17 compare responses for different listening groups when solo instrument source location is described. In both means and 95% confidence interval graphs, the centre of the vehicle (0) is represented by the reference line............
The following graphs (figure 4.D.18 - 4.D.21) illustrate means and 95% confidence intervals for positions of each item of programme material when described by all listeners. In the graphs, the centre of the vehicle (positioned at 0) is indicated by the reference line \( \_ \_ \_ \_ \), the left boundary of the vehicle is referred to by \( \_ \_ \_ \_ \_ \_ \) and the right vehicle boundary is denoted by \( \_ \_ \_ \_ \_ \_ \_ \_ \_ \).
Appendix 4.E U-GAL questionnaire responses

Questions posed:
1. Did you find it easy to use the user interface? (If not, why not?)
2. Did the provided shapes allow for the size of the sound that you heard? (If not, why not?)
3. Were there any spatial characteristics you heard but found yourself unable to describe using the provided shapes?
4. Were there any other characteristics that you would have liked to describe?
5. Anything else you would like to mention?

Listener responses:

<table>
<thead>
<tr>
<th>Question 1</th>
<th>Question 2</th>
<th>Question 3</th>
<th>Question 4</th>
<th>Question 5 (comments)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>-</td>
</tr>
<tr>
<td>2 NO: Problems moving shapes</td>
<td>NO: space bubble difficult to get to right size; maybe need more increments</td>
<td>NO</td>
<td>Height</td>
<td>-</td>
</tr>
<tr>
<td>3 YES</td>
<td>YES</td>
<td>NO</td>
<td>Height</td>
<td>Easier having carried out test previously. New users may need further explanation.</td>
</tr>
<tr>
<td>4 YES</td>
<td>YES</td>
<td>NO</td>
<td>Triangle shape</td>
<td>-</td>
</tr>
<tr>
<td>5 YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>Didn’t use the cloud shape at all because I felt like I was in small space throughout</td>
</tr>
<tr>
<td>6 YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>-</td>
</tr>
<tr>
<td>7 YES</td>
<td>YES</td>
<td>Sound may be louder on one part of region/ shape</td>
<td>Shape/location change as you hear same sample couple of times</td>
<td>-</td>
</tr>
<tr>
<td>8 YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>The system is a very comprehensive test. Very easy to describe a sound’s position and spaciousness</td>
</tr>
<tr>
<td>9 YES</td>
<td>YES</td>
<td>NO</td>
<td>Height of sound within vehicle</td>
<td>-</td>
</tr>
<tr>
<td>10 YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>Very Good - user friendly</td>
</tr>
<tr>
<td>12 YES</td>
<td>YES</td>
<td>NO</td>
<td>Liked to have heard music from behind me not just all from in front</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>---</td>
<td>-----</td>
<td>-----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>13</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>14</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>15</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>16</td>
<td>YES</td>
<td>NO: crescent shape would be more useful</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>18</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>19</td>
<td>YES</td>
<td>YES/NO: When resizing would have liked to change shape more</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>20</td>
<td>YES: but can't pick up letters under a shape</td>
<td>NO: Need to be able to rotate sound area</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>21</td>
<td>NO: Too many options</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>22</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
</tr>
</tbody>
</table>

Note: Comment sheets were not completed by listeners 11 and 17.
REFERENCES


N Ford

Doctoral Thesis

284


Nind, T. Multimedia in cars: The Use of Logic 7 Surround Processing as the Solution to the Challenge of Providing Surround Sound in Cars from all 2 channel and Encoded 5.1 Sources. *Audio Engineering Society 110th Convention*. May 2001, Preprint 5286


The graphical representation of listeners' auditory spatial experiences


N Ford

Doctoral Thesis

287