Phase as a Feature Extraction Tool for Audio Classification and Signal Localisation

Ioannis Paraskevas

Submitted for the Degree of Doctor of Philosophy from the University of Surrey

Centre for Vision, Speech and Signal Processing
School of Electronics and Physical Sciences
University of Surrey
Guildford, Surrey GU2 7XH, U.K.

2005

© Ioannis Paraskevas 2005
Abstract

The aim of this research is to demonstrate the significance of signal phase content in time localization issues in synthetic signals and in the extraction of appropriate features from acoustically similar audio recordings (non-synthetic signals) for audio classification purposes. Published work, relating to audio classification, tends to be focused on the discrimination of audio classes that are dissimilar acoustically. Consequently, a wide range of features, extracted from the audio recordings, has been appropriate for the classification task. In this research, the audio classification application involves audio recordings (digitized through the same pre-processing conditions) that are acoustically similar and hence, only a few features are appropriate, due to the similarity amongst the classes.

The difficulties in processing the phase spectrum of a signal have probably led previous researchers to avoid its investigation. In this research, the sources of these difficulties are studied and certain methods are employed to overcome them. Subsequently, the phase content of the signal has been found to be useful for various applications. The justification of this, is demonstrated via audio classification (non-synthetic signals) and time localization (synthetic signals) applications.

Summarizing, the original contributions, introduced based on this research work, are the ‘whitened’ Hartley spectrum and its short-time analysis, as well as the use of the Hartley phase cepstrum as a time localization tool and the use of phase related feature vectors for the audio classification application.
Acknowledgements

First, I would like to thank my supervisor, Dr. Edward Chilton, for his scientific guidance and patience throughout my work, as well as Professor Kittler for his financial support. I also owe my gratitude to my friends Dr. Costas Tzaras and Dr. Maria Faraklioti for their moral support.

I am indebted to my cousin Professor Constantina Skanavis for her encouragement and guidance in every aspect. Most importantly, I would like to thank my parents and dedicate this study to them.
## CONTENTS

1 Introduction 1  
   1.1 Aim of the research 1  
   1.2 Roadmap of the thesis 4  

2 Theoretical overview 7  
   2.1 Feature extraction from audio signals 10  
      2.1.1 Average zero-crossing rate 12  
      2.1.2 Volume 13  
      2.1.3 Linear prediction coefficients 14  
      2.1.4 Pitch 15  
      2.1.5 Cepstral coefficients and mel-frequency cepstral 16  
         coefficients  
      2.1.6 Magnitude spectrogram 18  
      2.1.7 Statistical and other features extracted from the 19  
         magnitude spectrum/spectrogram  
   2.2 Summary of the features extracted for audio classification 20
2.3 The phase information | 21
2.4 Summary of the theoretical overview | 24

3 Transforms and time-frequency distributions | 25
3.1 The Hartley transform | 26
  3.1.1 The Hartley/Fourier spectral relationship | 26
  3.1.2 The Discrete Hartley transform | 28
  3.1.3 Short-time analysis and the Hartley phase spectrum | 29
3.2 The homomorphic deconvolution process | 31
3.3 Time-frequency signal representation | 32
3.4 Quadratic time-frequency signal representation | 34
3.5 Visualisation of a time-frequency representation | 37
3.6 Summary | 38

4 Classification and other practical considerations | 41
4.1 The sampling process | 42
  4.1.1 The sampling process and aliasing | 42
  4.1.2 The sampling process and the DFT | 44
4.2 Windowing | 45
4.3 Evaluation of polynomial 'roots' | 47
4.4 Statistical measurements | 48
  4.4.1 Location statistics | 50
  4.4.2 Scale statistics | 51
  4.4.3 Shape and distribution statistics | 51
  4.4.4 Quantile statistics | 51
  4.4.5 Count statistics | 52
  4.4.6 Statistical feature vector | 52
4.5 Differential of the phase related spectra | 55
4.6 Distance metric classifiers | 56
4.7 Summary | 58
5 Phase analysis and preliminary results
5.1 Phase spectrum implementation based on the Fourier transform
5.1.1 Compensation of the ‘extrinsic’ discontinuities
5.1.2 Compensation of the ‘intrinsic’ discontinuities (via the DTFT)
5.1.3 Removal of the ‘intrinsic’ discontinuities (via the z-transform)
5.1.4 Summary of the methods for the compensation/removal of the ‘extrinsic’/‘intrinsic’ discontinuities
5.2 Experimental results of synthetic signals based on the DTFT approach
5.3 Phase spectrum implementation based on the Hartley transform
5.3.1 Hartley phase spectrum implementation via the DTHT
5.3.2 Hartley phase spectrum implementation via the z-transform
5.4 Experimental results of synthetic signals based on the DTHT approach
5.5 Step-by-step description of the phase related algorithms
5.6 Summary and conclusions

6 Experimental results of synthetic signals for the time localisation application
6.1 Experimental results of synthetic signals
6.1.1 Introduction to Fourier and Hartley phase cepstrums of pulses with noise
6.1.2 Fourier phase cepstrums of pulses with noise
6.1.3 Hartley phase cepstrums of pulses with noise
6.1.4 ‘Enhanced’ method
6.2 Behavior of the ‘whitened’ Hartley spectrum in a ‘simulated’ scenario
6.3 Behavior of the Fourier and the Hartley phase functions in a noisy environment 140
6.4 Summary and conclusions 143

7 Experimental results of non-synthetic signals for the audio classification application 145
7.1 Database 149
7.2 Preliminary classification results, signal representation selection, decision rule, cross-validation and significance test 151
    7.2.1 Signal representation selection 153
    7.2.2 Decision rule 155
    7.2.3 Cross-validation 158
    7.2.4 Significance (Matched-pairs t-test) 160
7.3 Classification results of the combinatorial scheme 162
7.4 Observations from the phase related spectrograms 164
    7.4.1 The use of phase ‘difference’ for feature extraction from the phase spectrograms 166
    7.4.2 Variation of the classification rate with respect to the compensation/removal of the discontinuities from the phase spectrograms 167
    7.4.3 Variation of the classification rate with respect to the sampling frequency of the phase spectrograms 169
    7.4.4 Comparison of the Hartley phase ‘difference’ with the Fourier phase ‘difference’ spectrograms in terms of classification 170
7.5 Summary and conclusions 172

8 Conclusions and future plans 173
8.1 Conclusions of the research 173
8.2 Future plans of the research 175
    8.2.1 ‘Root’ finding method of high-order polynomials 175
<table>
<thead>
<tr>
<th>Contents</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>8.2.2 Dimensionality reduction of the phase related spectrograms</td>
<td>175</td>
</tr>
<tr>
<td>8.2.3 Applications of the Hartley phase spectrum</td>
<td>176</td>
</tr>
<tr>
<td><strong>Appendices</strong></td>
<td></td>
</tr>
<tr>
<td>Appendix A1</td>
<td>177</td>
</tr>
<tr>
<td>Appendix A2</td>
<td>180</td>
</tr>
<tr>
<td><strong>Bibliography</strong></td>
<td>185</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 1.1 Pattern recognition system model 2

Figure 2.1: Triangular band-pass filter set 18

Figure 3.1: Summary of the homomorphic deconvolution process 31
Figure 3.2: Fourier magnitude spectrogram of a referee whistle 38

Figure 5.1a: Multiple excitation signal 65
Figure 5.1b: Magnitude spectrum of the multiple excitation signal 65
Figure 5.1c: Real part of the multiple pulses 65
Figure 5.1d: Imaginary part of the multiple pulses 65
Figure 5.2a: Fourier phase spectrum – removal of the ‘extrinsic’
discontinuities 66
Figure 5.2b: Fourier phase spectrum – removal of the ‘extrinsic’ and
the ‘intrinsic’ discontinuities 66
Figure 5.3: Calculation of phase response of a ‘zero’ in the z-plane 68
Figure 5.4: ‘Zeros’ located in the z-plane (of a single 256 samples long frame from class (vii))

Figure 5.5: ‘Zeros’ located in the z-plane (of a single – same as in figure 5.4 - 256 samples long frame from class (vii), after the application of the exclusion ‘ring’ with width 0.004)

Figure 5.6: ‘Zeros’ located in the z-plane of an exponentially decaying sine wave (located from 251 to 280)

Figure 5.7: ‘Zeros’ located in the z-plane of a rectangular pulse (located from 251 to 280, figure 5.11a)

Figure 5.8: Summary of the homomorphic deconvolution process

Figure 5.9: Stages of the non-linear part of the homomorphic deconvolution process applied to the Fourier case

Figure 5.10a: Time domain representation of a rectangular pulse

Figure 5.10b: Frequency (Fourier phase spectrum) domain representation of a rectangular pulse

Figure 5.10c: Cepstral (Fourier phase cepstrum) domain representation of a rectangular pulse

Figure 5.11a: Time domain representation of a rectangular pulse

Figure 5.11b: Frequency (Fourier phase spectrum) domain representation of a rectangular pulse

Figure 5.11c: Cepstral (Fourier phase cepstrum) domain representation of a rectangular pulse

Figure 5.12a: Time domain representation of a rectangular pulse

Figure 5.12b: Frequency (Fourier phase spectrum) domain representation of a rectangular pulse

Figure 5.12c: Cepstral (Fourier phase cepstrum) domain representation of a rectangular pulse

Figure 5.13: Phase response of a ‘zero’ in the z-plane

Figure 5.14: Stages of the non-linear part of the homomorphic deconvolution process applied to the Hartley case

List of Figures

70

70

73

73

74

74

79

79

79

80

80

80

81

81

81

81

84

86
List of Figures

Figure 5.15a: Time domain representation of a rectangular pulse 93
Figure 5.15b: Frequency (Hartley phase spectrum) domain representation of a rectangular pulse 93
Figure 5.15c: Cepstral (Hartley phase cepstrum) domain representation of a rectangular pulse 93
Figure 5.16a: Time domain representation of a rectangular pulse 94
Figure 5.16b: Frequency (Hartley phase spectrum) domain representation of a rectangular pulse 94
Figure 5.16c: Cepstral (Hartley phase cepstrum) domain representation of a rectangular pulse 94
Figure 5.17a: Time domain representation of a rectangular pulse 95
Figure 5.17b: Frequency (Hartley phase spectrum) domain representation of a rectangular pulse 95
Figure 5.17c: Cepstral (Hartley phase cepstrum) domain representation of a rectangular pulse 95
Figure 5.18a: Frequency (Hartley phase spectrum) domain representation of a rectangular pulse – ‘intrinsic’ discontinuities not removed 96
Figure 5.18b: Frequency (Hartley phase spectrum) domain representation of a rectangular pulse – ‘intrinsic’ discontinuities removed 96
Figure 5.18c: Cepstral (Hartley phase cepstrum) domain representation of a rectangular pulse – ‘intrinsic’ discontinuities removed 96

Figure 6.1a: Time domain of a rectangular pulse with 10.2 dB SNR 121
Figure 6.1b: Cepstral domain – ‘intrinsic’ discontinuities not compensated 121
Figure 6.1c: Cepstral domain – ‘intrinsic’ discontinuities compensated 121
Figure 6.2a: Time domain of five periodically spaced exponentially decaying sine waves 128
Figure 6.2b: ‘Whitened’ Hartley spectrum – ‘intrinsic’ discontinuities not compensated 128
Figure 6.2c: ‘Whitened’ Hartley cepstrum
Figure 6.3a: Time domain of five periodically spaced exponentially
decaying sine waves
Figure 6.3b: ‘Whitened’ Hartley spectrum – ‘intrinsic’ discontinuities
compensated
Figure 6.3c: ‘Whitened’ Hartley cepstrum
Figure 6.4a: Time domain of five periodically spaced exponentially
decaying sine waves
Figure 6.4b: ‘Whitened’ Hartley spectrum – ‘intrinsic’ discontinuities
compensated - (‘enhanced’) method
Figure 6.4c: ‘Whitened’ Hartley cepstrum
Figure 6.5a: ‘Artificial’ ‘whitened’ Hartley spectrum
Figure 6.5b: Subtraction of the ‘whitened’ from the ‘artificial’
‘whitened’ Hartley spectrum
Figure 6.5c: Difference of the subtraction of the ‘whitened’ from
the ‘artificial’ ‘whitened’ Hartley spectrum - (figures 6.3b, 6.5a)
Figure 6.5d: Difference of the subtraction of the ‘whitened’ from
the ‘artificial’ ‘whitened’ Hartley spectrum - (figures 6.4b, 6.5a)
Figure 6.6a: Time domain of three periodically spaced exponentially
decaying sine waves
Figure 6.6b: ‘Whitened’ Hartley spectrum - ‘intrinsic’ discontinuities
not compensated
Figure 6.6c: ‘Whitened’ Hartley cepstrum
Figure 6.7a: Time domain of three periodically spaced exponentially
decaying sine waves
Figure 6.7b: ‘Whitened’ Hartley spectrum – ‘intrinsic’ discontinuities
compensated
Figure 6.7c: ‘Whitened’ Hartley cepstrum
Figure 6.8a: Time domain of three periodically spaced exponentially
decaying sine waves
List of Figures

Figure 6.8b: 'Whitened' Hartley spectrum – 'intrinsic' discontinuities compensated - ('enhanced') method 135
Figure 6.8c: 'Whitened' Hartley cepstrum 135
Figure 6.9a: 'Artificial' 'whitened' Hartley spectrum 136
Figure 6.9b: Subtraction of the 'whitened' from the 'artificial' 'whitened' Hartley spectrum 136
Figure 6.9c: Difference of the subtraction of the 'whitened' from the 'artificial' 'whitened' Hartley spectrum - (figures 6.7b, 6.9a) 136
Figure 6.9d: Difference of the subtraction of the 'whitened' from the 'artificial' 'whitened' Hartley spectrum - (figures 6.8b, 6.9a) 136
Figure 6.10a: Time domain of the 'simulated' scenario 139
Figure 6.10b: Cepstral domain of the 'simulated' scenario 139
Figure 6.11a: Probability density function (pdf) of noise in the Hartley phase spectrum 142
Figure 6.11b: Probability density function (pdf) of noise in the Fourier phase spectrum 142

Figure 7.1: Fourier magnitude spectrogram of an audio recording from class (vi) 150
Figure 7.2: Fourier magnitude spectrogram of an audio recording from class (x) 150

Figure 6.1A: Coordinate system, $\frac{\pi}{4}$ radians tilted with respect to the conventional Cartesian system 181
LIST OF TABLES

Table 5.1: Amplitude of the 0th cepstral value in the Fourier phase cepstral domain for the exponentially decaying sine wave pulse and for the rectangular pulse  
Table 5.2: Location of the maximum amplitude in the Hartley phase cepstral domain for the exponentially decaying sine wave pulse and for the rectangular pulse  
Table 6.1: Amplitude of the 0th cepstral value in the Fourier phase cepstral domain for the exponentially decaying sine wave pulse  
Table 6.2: Amplitude of the 0th cepstral value in the Fourier phase cepstral domain for the rectangular pulse  
Table 6.3: Location of the maximum amplitude in the 'whitened' Fourier cepstral domain for the exponentially decaying sine wave pulse  
Table 6.4: Location of the maximum amplitude in the 'whitened' Fourier cepstral domain for the rectangular pulse
Table 6.5: Location of the maximum amplitude in the Hartley phase cepstral domain for the exponentially decaying sine wave pulse 122
Table 6.6: Location of the maximum amplitude in the Hartley phase cepstral domain for the rectangular pulse 123

Table 7.1: Database of gunshots 150
Table 7.2: Classification scores in % over all the ten classes 152
Table 7.3: Classification scores in % (for classes 2, 4, 5 & 10 and for all 10 classes) derived from the combinatory scheme 163
1.1 Aim of the research

The aim of this research is to demonstrate the significance of signal phase content in time localization issues in synthetic signals and in the extraction of appropriate features from acoustically similar audio recordings (non-synthetic signals) for audio classification purposes. The published literature that is related to the audio classification application is focused on the discrimination of audio recordings that although they belong to the same group in terms of content (e.g., sports sounds), are dissimilar acoustically. Acoustically similar audio classes convey similar spectral properties whereas, acoustically dissimilar audio classes convey dissimilar spectral characteristics, although they may belong to the same category in terms of content (e.g., sports sounds). Therefore, due to their acoustical dissimilarity, a wide range of feature extraction methods may be appropriate for the pattern recognition task. Features, such as LPC (Linear Prediction Coefficients) or cepstral coefficients, have been applied to speech recognition systems and were justified from a model of the human speech production structure. However, such a model may not
be appropriate for the general audio classification problems other than speech. Moreover, due to the dissimilarity among the audio classes, many of the features extracted are based on apriori information related to the spectral shape of the audio recording or based on apriori information related to time-dependent features (e.g., time duration). However, the heuristic nature of the time-dependent features makes them inappropriate for cases where the audio classes are acoustically similar.

Hence, the system model can be divided, as for any other pattern recognition system model, into two stages: the feature extraction and the classification (as shown in figure 1.1).

![Figure 1.1 Pattern recognition system model](image)

In the feature extraction part, features, that encapsulate the information content, are extracted from each audio recording. In the decision making part, the audio recordings are classified to certain classes according to the feature vectors formed in the feature extraction stage.

It is important to mention that the center of attention of the research in the audio classification application is the feature extraction part of the pattern recognition process. Hence, the statistical features, extracted from the time-frequency representations of the audio recordings, are presented to a distance-metric classifier.

The database used for this work consists of ten different classes of gunshots that are acoustically similar. Consequently, due to this similarity, it is not possible to use apriori information and thus, the feature extraction part is based entirely on non-parametric
frequency-based features. However, although the magnitude spectrum is a very useful source of information on its own it is inadequate, as it needs additional feature streams to reinforce the classification process or even as a substitute, when it fails to classify certain classes.

The difficulties in processing the phase spectrum of non-synthetic signals have probably led previous researchers to avoid its investigation [104]. Consequently, until recently, the published literature related to the phase content of one dimensional signals (1D), is limited to parametric methods of phase estimation specialized for synthetic signals. However, in this research, novel phase analysis techniques have been introduced (chapter five), which show that in the case where the difficulties in processing the phase spectrum are overcome then useful information can be extracted. Hence, as shown from the experimental results obtained for the classification application, the phase information combined with the magnitude information increases the classification rate significantly, compared to the case where only the magnitude information is used.

In order to explain every step of the process followed so as to overcome the difficulties appearing in the phase spectrum analysis, the application of time localization/time delay measurement for synthetic signals (rectangular, exponential pulses, etc.), is presented. The aim of the time localization/time delay application is to differentiate the phase from the magnitude information so as to investigate its shape, behavior and other properties in a noise-free/noisy environment. Similarly to the audio classification application, the experimental results of the time localization of a signal application show that the Hartley phase spectrum, unlike the Fourier phase spectrum, encapsulates the phase content of the signal more efficiently, without suffering from ‘wrapping’ ambiguities, even in the case where the environment is noisy.
1.2 Roadmap of the thesis

The thesis is organized in three divisions: the first division includes chapter two, the second division includes chapters three and four while, the third division includes chapters five, six and seven. The second chapter presents the literature review related to the audio classification application and the phase analysis of one dimensional (1D) signals. The third chapter, amongst others, presents a description of the Hartley transform used for the time localization of a signal and the audio classification applications while, chapter four presents some practical issues that are related to the experimental set up of the audio classification application. The major part of the thesis though is in the third division (chapters five to seven). Hence, in chapter five the theory related to the removal/compensation of the discontinuities from the phase function is addressed. In the same chapter, the Fourier phase cepstrum and the Hartley phase cepstrum are introduced, together with examples related to the time localization of a signal application in a noise-free environment. In the sixth chapter the comparison of the Fourier phase cepstrum with the Hartley phase cepstrum is extended for the case where the environment is noisy. In the seventh chapter, the audio classification application is presented. Finally, in chapter eight, the conclusions and possible future developments of the research, are presented.

Specifically, the original contributions introduced based on this research work are:

a) the 'whitened' Hartley spectrum (Hartley phase spectrum) in chapter three
b) the short-time analysis of the 'whitened' Hartley spectrum in chapter three
c) the Hartley phase cepstrum as a time localization tool in chapters five and six
d) the use of phase related feature vectors for the audio classification application in chapter seven
Part of the conclusions/observations derived from the research related to the audio classification application is published in the following conference and journal papers:

Title: ‘Fine classification for acoustic images’
Authors: Ioannis Paraskevas, Edward Chilton
April 2003, International Signal Processing Conference, GSPx, Dallas, Texas, USA

Title: ‘Audio fine classification using the statistical analysis of acoustic images’
Authors: Edward Chilton, Ioannis Paraskevas
April 2003, Acoustical Society of America, Nashville, USA

Title: ‘Audio classification for retrieval from multimedia databases’
Authors: Ioannis Paraskevas, Edward Chilton
July 2003, 4th EURASIP Conference focused on Video/Image Processing and Multimedia Communications, Zagreb, Croatia

Title: ‘Phase as an assistive feature vector for audio classification’
Authors: Ioannis Paraskevas, Edward Chilton
March 2004, Institute of Acoustics, Spring Conference, Southampton, UK

Title: ‘Acoustic classification using time-frequency distributions’
Authors: Hossein Marvi, Ioannis Paraskevas, Edward Chilton
March 2004, Institute of Acoustics, Spring Conference, Southampton, UK

Title: ‘Combination of Magnitude and Phase Statistical Features for Audio Classification’
Authors: Ioannis Paraskevas, Edward Chilton
Acoustics Research Letters Online (ARLO), Acoustical Society of America, July 2004
CHAPTER 2
THEORETICAL OVERVIEW

Introduction

This chapter presents an overview of the most recent and relevant literature in feature extraction for the application of audio classification, together with a discussion of some of the theoretical aspects of this work. The research in feature extraction from audio sounds [23] is relatively new, when compared to the research in feature extraction from speech. Consequently, most of the feature extraction methods, for audio classification, are based on certain assumptions that hold for the case of speech feature extraction. However, the human vocal tract model assumes a linear filter (vocal tract) excited by an independent source of acoustic energy (pulsed or random noise). The broader class of acoustic signals cannot be modeled in this way and hence, the assumptions adopted for feature extraction from speech do not necessarily hold for the case of audio.

In this part of the introduction, an overview of the literature, related to information retrieval from audio-visual databases using audio as a feature vector is provided. According to [4] and [5], the audio signal is segmented and then classified into the
following categories: speech, music, environmental sounds and silence [1]. Some of the features used are the energy of the signal, the zero-crossings, the fundamental frequency and periodicity. In [3], another method is proposed for the same process, in which the feature extraction is based on the volume distribution, pitch contour and other frequency related features. Hence, the audio signals are categorized in the following groups: news, weather reports, advertisements, football and basketball games. However, there appears to be no direct relationship between the chosen features and the content-based classes that is attempted to be identified. The feature extraction method proposed in [2], is based on mel-scale cepstral coefficients, while in [6] another method for content-based audio classification is proposed, where a combination of perceptual [9] and cepstral features is used. In [8] and [55] the audio recognition research is extended by using combinations of features. These included mel-scale cepstral coefficients, pitch value [12] and zero crossing rate. Finally, some of the features, which the audio recognition tool of MPEG-7 uses, are the fundamental frequency, the frequency centroid, the energy and the harmonicity [16] [17] [52]. In this work, the features used are statistical measurements (subsection 4.4.6, chapter four and subsection 2.1.7) extracted from the spectrograms of the acoustic events to be classified (chapter seven).

Up to now, most of the research related to audio classification and retrieval, is focused on the discrimination between audio classes that are acoustically dissimilar. For example, sport sounds [18], [19], [20], [22] (i.e. tennis hit [15], [21] swimming bell, whistle referee etc.) belong to the same family, however, are acoustically dissimilar. A research area that aims to classify audio classes that are acoustically dissimilar is the computational auditory scene analysis [116], [117]. Consequently, speech-related or even temporal-related feature extraction techniques, can provide high classification scores, due to the dissimilarity among the classes of the audio utterances [51]. Moreover, for most of the cases of event detection and retrieval from multimedia databases, audio is employed as an assistive feature vector for classification [11], and hence is combined with image-related feature vectors [54]. One scientific area that, sometimes, employs audio as the principal feature vector for classification, and also aims to discriminate classes of utterances that are acoustically similar, is zoology (discrimination of bat species using audio as the
feature vector for classification [53]). Finally, two other scientific areas that aim to classify audio classes that sometimes maybe acoustically similar are musical instrument classification [113] and musical genre classification of audio signals [114], respectively.

There are two reservations to the existing feature extraction methods, that may limit the effectiveness of these techniques. The first, is that features such as LPC or cepstral coefficients were developed for speech recognition systems and their justification was based on a model of the human speech production system. However, the aforementioned as well as other speech-based observations are not necessarily applicable to the whole range of audio. The second reservation is that the extracted features represent the magnitude spectrum of the acoustic unit only and, consequently, phase is not represented. Hence, as it will be discussed in section 2.3 and chapter seven, one aspect of our work is the extraction of features from the phase spectrum of the acoustic signal to assist in the classification process.

In the work that will be presented in the following chapters, an alternative approach is described for the extraction of features for audio classification. In the experimental part (chapter seven) the audio utterances used, belong to different classes of gunshots and thus, the audio samples of the database are acoustically similar ('fine' audio classification). In the proposed method, the features extracted from a spectral (frequency domain) representation of the audio signal, include both magnitude and phase information. The approach firstly uses a suitable transform (such as the Fourier/Hartley) to generate a time-frequency surface via a sliding short-time window (section 3.3, chapter three). A statistical analysis is then applied to the time-frequency surfaces, so as to reduce the information dimensionality, so that the content of the signal may be presented to the classifier in a compact manner. Two of the research publications that propose a similar feature extraction approach, are [13] and [14]. Thus, [14] proposes the extraction of features from the magnitude spectrogram, whereas [13] proposes the extraction of statistical features from the time and wavelet transform domains. However, unlike the work discussed here, these researchers do not use the phase content of the acoustic events.
Note that a description of the classification part of the pattern recognition process will not be included in the literature review chapter, as the aim of the research is focused on the feature extraction from the time-frequency distributions formed by the audio utterances. However, in order to test the performance of different sets of feature extraction these must be tested on a classifier and so the issue of classifier design is of some relevance to the research. The classification issue will be discussed in chapters four (classification and other practical considerations) and seven (audio classification application).

In section 2.1, there will be a description of some of the most popular, time-based, frequency-based and time-frequency-based audio feature extraction techniques. Moreover, due to the fact that all the features used for audio classification are speech oriented, the examples provided will be related to speech. In section 2.3, there will be an overview of the literature related to the phase analysis of one dimensional signals and in section 2.4, a summary of this chapter is provided.

2.1 Feature extraction from audio signals

In this section, the various features that can be extracted from an audio signal are discussed. Before mentioning the mathematical expressions of the most popular of them, an overview of the categories, into which they can be divided, is provided.

There are many ways to categorize the features that can be extracted from an audio signal [10]. One way is to categorize them based upon the domain from which they are extracted. Hence, there are features extracted from the time domain (e.g., energy, zero-crossing rate (ZCR), volume, volume dynamic range (VDR), etc.) and features extracted from the frequency domain (e.g., brightness, bandwidth, pitch, etc.). Moreover, features or coefficients can be extracted from time-frequency representations of the signal (e.g., magnitude and phase spectrograms) and coefficient representations of the signal (e.g., cepstral coefficients, mel-scale frequency cepstral coefficients (MFCC), linear prediction coefficients (LPC), etc.), respectively. Also, the features extracted from an audio signal can be divided into the physical features (i.e. features that convey audio properties related
to physical quantities, e.g. fundamental frequency, ZCR, energy, etc.) and the perceptual
features (i.e. features that convey audio properties related to the way humans perceive
sound (e.g., rhythm, timbre, loudness, pitch, etc.). The features found to be most useful
for this work are those extracted from the time-frequency signal representations.

In the remaining part of this section, there will be an overview of the most popular
features for 'coarse' (discrimination between audio classes that are not acoustically
similar) audio classification. The feature extraction methods that are related to 'coarse'
audio classification are presented in order to demonstrate the disadvantages that some of
them convey and consequently, select the most appropriate ones employed in our work.
Hence the features that will be presented are organized as follows: the first category of
features are related to the time domain of the signal, whereas the second, includes the
features that are related to the signal's frequency domain. The features that belong to the
time domain are: i) Average zero-crossing rate ii) Volume and iii) Linear Prediction
Coefficients. The features that belong to the frequency domain are:

i) Pitch, ii) Cepstral coefficients and mel-frequency cepstral coefficients iii) Frequency
centroid and iv) Bandwidth. Finally, the magnitude spectrogram is a time-frequency
signal representation from which additional features can be extracted. Hence, as it will be
discussed later, some of the features that belong to the frequency domain, such as the
frequency centroid and the bandwidth, can be extracted from the aforementioned time-
frequency signal representation.
2.1.1 Average Zero-Crossing Rate

Definition:

In the context of discrete time signals, a zero-crossing is said to occur, if successive samples have different signs [107]. The zero-crossing rate (ZCR) is defined as:

\[
ZCR = \frac{1}{2(N-1)} \left( \sum_{n=2}^{N} | \text{sgn}[s(n)] - \text{sgn}[s(n-1)] | \right) \tag{2.1},
\]

where, \( n \) is the discrete time index and \( s(n) \) is the discrete time signal of length \( N \),

\[
\text{sgn}[s(n)] = \begin{cases} 
1, & s(n) \geq 0 \\
-1, & s(n) < 0 
\end{cases} \tag{2.2}
\]

and \(| |\) denotes absolute value.

The zero-crossing rate [107] is used as an indication of discrimination between speech and silence. Generally, speech components have much higher zero-crossing rate compared to silence.

In order to distinguish between other kinds of audible sounds, based on the zero-crossing rate criterion, it is possible to use other characteristics of the zero-crossing rate, such as regularity, periodicity, stability and amplitude range [4]. A detailed discussion of these additional features is beyond the scope of this thesis.
2.1.2 Volume

Definition:

The volume (also referred as loudness or root mean square (RMS) value) of a signal is defined as:

\[ v = \sqrt{\frac{1}{N} \sum_{n=1}^{N} s^2(n)} \]  

(2.3)

where, \( s(n) \) is the discrete time signal of length \( N \).

It has to be noted that the terms volume and loudness are not appropriate scientifically. However, both terms are often used, in the published literature, instead of the scientifically appropriate RMS value term.

The mean and the standard deviation of the volume (calculated over a certain number of frames) of an audio signal can be used as descriptors [3]. An audio frame can be characterized as silent or not, based on the comparison of its volume with a threshold determined by the volume distribution of the entire audio clip. After the silence is detected, it is possible to calculate the silence ratio, which is defined as the ratio of the silence interval over the length of the entire audio clip.

Volume is a helpful tool in the discrimination between unvoiced speech and silence. The characteristics of unvoiced speech and the noise remaining on the signal during periods of silence (absence of speech) maybe very similar. However, one assumes that the volume of the signal during unvoiced speech is higher when compared to periods of silence. A feature derived from the volume is the Volume Dynamic Range (VDR), defined as: \( (\text{max}(v) - \text{min}(v))/\text{max}(v) \), where the \( \text{max}(v) \) and the \( \text{min}(v) \) are the maximum and the minimum volumes within an audio clip, respectively.
2.1.3 Linear Prediction Coefficients

Linear prediction is a time domain analysis technique [48] that is widely applied to speech analysis and coding. The assumption, which this technique is based on, is that the present speech sample can be approximated as a linear combination of its past samples.

Hence,

\[ s(n) = a_0 + a_1 s(n-1) + a_2 s(n-2) + \ldots + a_p s(n-p) = \sum_{k=1}^{p} a_k s(n-k) \quad (2.4) \]

Thus, in speech, the relationship between the speech samples \( s(n) \) at the output of the vocal tract filter and the input to the filter (i.e., excitation) is given by the following equation:

\[ s(n) = \sum_{k=1}^{p} a_k s(n-k) + Gu(n) \quad (2.5), \]

where \( G \) is the gain, \( a_k \) are the filter coefficients and \( u(n) \) is the normalized excitation.

The same equation can be expressed, in the z-domain as:

\[ S(z) = S(z) \sum_{k=1}^{p} a_k z^{-k} + GU(z) \quad (2.6) \]

Thus, from equation (2.6), the vocal tract transfer function, \( H(z) \), can be written as:

\[ H(z) = \frac{S(z)}{GU(z)} = \frac{1}{1 - \sum_{k=1}^{p} a_k z^{-k}} = \frac{1}{A(z)} \quad (2.7) \]

Feature vectors formed from LPCs do not distribute themselves in the feature vector space in a manner suitable for direct classification. Thus, the LPC have to be transformed
to a new set of parameters, such as cepstral coefficients or line spectral pairs, so as to perform better in classification related applications [57].

Many researchers are using the LPC, or coefficients based on the LPC, for audio classification problems. However, Linear Prediction analysis is dependent on a source filter model appropriate for speech analysis but not for general audio.

2.1.4 Pitch

Definition:

Pitch [7] is a perceptual feature of the audio signal which amongst others, depends on the fundamental frequency of the audio waveform. Pitch information can be extracted by using either temporal or frequency analysis.

The temporal analysis method is based on the computation of the Autocorrelation Function, \( R_n(l) \) [46] or AMDF (Average Magnitude Difference Function), \( A_n(l) \) [47]. Thus,

\[
R_n(l) = \sum_{n=1}^{N-1} s(n)s(n+1) \quad \text{and} \quad A_n(l) = \sum_{n=1}^{N-l} |s(n+1)-s(n)|
\]

where \( s(n) \) is the discrete time audio signal and \( l \) is the shift.

Based on the frequency analysis method, the pitch can be determined from the periodic structure in the magnitude spectrum of the Fourier transform, or the cepstral coefficients [58] of an audio frame. Generally, well-defined pitch characterizes only speech and harmonic music, but it can also be a feature of other audible waveforms (bats, dolphins etc.). In speech, the pitch frequency depends basically on the gender of the speaker. In music signals, it depends on the strongest note being played. There are three different features that are used in order to estimate the pitch variation: the standard deviation of the pitch (indicates the variation of the value of pitch over the utterance), the smooth pitch ratio and the non-pitch ratio. The smooth pitch ratio feature estimates the percentage of
music or voiced speech that exists within a clip, since only music and voiced speech are characterized by smooth pitch. On the contrary, the non-pitch ratio estimates the percentage of unvoiced speech or noise within an audio clip, since these two have no pitch. Furthermore, a close relative of pitch is the feature of harmonicity, which measures the strength of pitch perception [12].

2.1.5 Cepstral coefficients and mel-frequency cepstral coefficients

Cepstral coefficients [108] [109] and mel-frequency cepstral coefficients (MFCCs) [110] are very popular in speech/speaker feature extraction [56]. The selected coefficients aim to represent the original spectrum in a compressed manner. The difference between these two compression techniques (i.e. cepstral coefficients and mel-frequency cepstral coefficients) is that the MFCC approach is based on the property of a non-linear scale of frequencies, similar to that of the human hearing system [7].

Speech is assumed to be the convolution of an excitation signal with the impulse response of the vocal tract filter, i.e.,

\[ s(n) = v(n) * e(n) \] (2.9)

where, \( s(n) \) is the discrete time speech signal, \( v(n) \) is the discrete time impulse response of the vocal tract filter, \( e(n) \) is the discrete time excitation signal and the asterisk symbol (*) denotes convolution. The cepstral analysis technique attempts to deconvolve the excitation signal from the vocal tract transfer function, without making some of the assumptions that are necessary for linear prediction [8] (section 2.1.3). The \( v(n) \) term (i.e. the impulse response of the vocal tract filter), corresponds to the shape of the vocal tract, the lip radiation and the glottal pulse, that convey linguistic information and which is distinct for each speaker. The \( e(n) \) term (i.e., excitation signal) may not convey such important information for the application of speech recognition and thus, the aim is to separate the excitation signal from the vocal tract transfer function and preserve only the information related to the latter.
Chapter 2. Theoretical overview

The real cepstrum is defined as:

\[ c(t) = IDTFT[\log|S(\kappa)|] \] (2.10)

where, \( c(t) \) represents the cepstral function, \( IDTFT \), the Inverse Discrete-Time Fourier Transform, \( \kappa \), the discrete frequency index and \( |S(\kappa)| \) the magnitude spectrum of the discrete time signal \( s(n) \). From equation (2.10),

\[ c(T) = ZTFT[\log|S(\kappa)|] = IDTFT[\log|E(\kappa)|] + IDTFT[\log|V(\kappa)|] \] (2.11)

where \( |E(\kappa)| \) and \( |V(\kappa)| \) are the discrete magnitude spectra of the discrete time signals \( e(n) \) and \( v(n) \), respectively. Equation (2.11) states that the convolution between the \( v(n) \) and \( e(n) \) terms, in the discrete time domain, is mapped to an addition in the cepstral domain. Moreover, for speech signals, these two logarithmic terms are well separated in the cepstral domain. Hence, most of the speech researchers are selecting the coefficients that correspond to the \( IDTFT[\log|V(\kappa)|] \) term only, by applying a window (liftering operation).

The mel-scale frequency cepstral coefficients (MFCCs) form a group of coefficients that is closely related to the cepstral coefficients. The MFCCs are related to the physiological structure of the human ear (i.e., variation of critical bandwidths with frequency). Thus, the mel-scale is linearly distributed in the lower frequencies whereas, it is logarithmically distributed in the higher ones [105] [110]. The mathematical relationship between the linear and the mel scale frequencies is expressed as:

\[ M(f) = f \quad \text{for} \quad f \leq 1kHz \text{ and} \]

\[ M(f) = 1127 \log(1 + \frac{f}{700}) \quad \text{for} \quad f > 1kHz \] (2.12)
where, $M$ and $f$ correspond to the mel and linear frequencies, respectively. The mel-scale frequency coefficients can be obtained by simulating critical-band filtering with a set of triangular band-pass filters, as in figure 2.1.

![Triangular band-pass filter set](image)

**Figure 2.1: Triangular band-pass filter set (taken from [25])**

The filters are spaced linearly in the frequency range from 0 to 1000Hz. For frequencies higher than the 1000Hz, the center frequency of each filter is given by:

$$f_{rel} = 1.148 f_i$$  \hspace{1cm} (2.13) where the initial frequency ($f_i$), is 1kHz.

Cepstral and mel-scale cepstral coefficients have also been used in audio recognition. The issue that arises is whether such coefficients are appropriate for general audio recognition where in most cases the signal is not the product of a source filter model. However, it is the significant compression properties of the cepstrum that make it useful for general audio recognition tasks, since an estimation of a signal spectrum can be obtained from just a few low order cepstral coefficients.

2.1.6 Magnitude spectrogram

**Definition:**

The discrete magnitude spectrogram is a matrix that presents the evolution of the magnitude spectrum content of a signal, across time. Initially the signal is divided into frames, windowed and transformed to the frequency domain (Short-Time Fourier
Transform (STFT)). Then, the magnitude spectrum of each windowed frame is evaluated. Consequently, a matrix is formed. The first row of the matrix corresponds to the magnitude spectrum of the first time frame while, the last row of the matrix, to the magnitude spectrum of the last time frame. More details related to the time-frequency signal representations, are provided in chapter three, sections 3.3 to 3.5.

The major disadvantage of the use of the magnitude spectrogram, as a feature extraction tool, is its high dimensionality. Hence, peak detection, statistical measurements or other methods have to be applied to reduce its dimensionality. The peak track in the magnitude spectrogram of an audio signal often reveals important characteristics of the sound [5].

2.1.7 Statistical and other features extracted from the magnitude spectrum/spectrogram

In this subsection, some features that can be extracted from the magnitude spectrum/spectrogram, will be mentioned. Hence, the frequency centroid (FC) and the bandwidth (BW), of an audio frame, are defined as:

\[
FC = \frac{\sum_{\kappa=1}^{M} \kappa |S(\kappa)|}{\sum_{\kappa=1}^{M} |S(\kappa)|} \tag{2.14}
\]

\[
BW = \frac{\sum_{\kappa=1}^{M} (\kappa - FC)^2 |S(\kappa)|}{\sum_{\kappa=1}^{M} |S(\kappa)|} \tag{2.15}
\]

respectively, where, \( \kappa \) is the frequency index and \(|S(\kappa)|\) is the magnitude spectrum (\(M\) frequency points) of the audio signal. Note that, the FC is related to the human sensation of the brightness of a sound [7] and hence, sometimes in literature the latter term is used.

In our work, statistical features are extracted from the magnitude spectrogram and therefore, a more detailed description about statistical feature extraction from time-frequency signal representations will be provided in chapter four, section 4.4.
Moreover, the values of the aforementioned features (subsections 2.1.1 to 2.1.7) are derived with respect to a single audio frame. In the case where an audio clip is divided into smaller frames, the same features can be extracted from every single frame. The values of the features evaluated from each frame, can be compared with the values of the features from the rest of the frames, in order to extract additional information related to the signal. Moreover, further processing can be applied to the calculation of some features (volume, ZCR, etc.), derived from a certain number of frames, such as the estimation of their first or second derivative [8]. Consequently, even more features can be extracted, from the statistical analysis of the derivatives of these features. However note that, there are cases of information overlapping, in the case where features of similar content are extracted.

2.2 Summary of the features extracted for audio classification

In summary, all the aforementioned audio features have been applied to ‘coarse’ audio classification [7]. However, some of them may not be appropriate for the application of ‘fine’ (discrimination between audio classes that acoustically are similar) audio classification. For example, for the case of ‘goal’ detection from a database of football sounds, the volume of a ‘goal cheering’ frame is significantly higher compared to a ‘non-goal cheering’ frame. Thus, the volume conveys information related to the amplitude of the signal rather than information related to the structure of the signal. However, the database used for this research (chapter seven), consisted of different classes of gunshots (‘fine’ audio classification) and thus, volume is not a useful feature due to the similarity of the amplitude envelope between the utterances. Moreover, linear prediction, cepstral and mel-scale frequency cepstral coefficients are useful for speech related applications, because they are based on certain assumptions that are related to the vocal tract model. However, the same assumptions do not hold for the case of general audio utterances. Pitch is a feature that can discriminate classes of audio with different harmonicity. Audio utterances, that are acoustically similar (e.g. bats, gunshots etc.), are characterized by similar degrees of harmonicity and consequently, pitch is not a useful feature for ‘fine’ audio classification. Finally, the ZCR is a feature that encapsulates the information
related to the amount of noise an audio signal conveys. However, the database used (classes of gunshots) in our work consisted of sounds that conveyed similar amounts of noise and thus the ZCR is not a useful feature for class discrimination. Therefore, volume, LPCs, Cepstral Coefficients, MFCCs, pitch and ZCR are not always appropriate for 'fine' audio classification.

Therefore, from the features analyzed in section 2.1, the statistical measurements extracted from the magnitude spectrogram, are the most appropriate for the application of 'fine' audio classification because they are not based on speech related assumptions (such as the LPCs, the Cepstral coefficients and the MFCCs). Also, they encapsulate information that is not limited to the degree of harmonicity (such as the pitch), the amount of noise (such as the ZCR) or the amplitude envelope (such as the volume).

Finally, it is important to underline the issue of feature selection. Each of the features that are introduced to the classifier should encapsulate distinct information, so as to avoid overlapping (i.e., two or more features conveying similar information) [24]. Feature vectors of increased dimensionality may cause 'confusion' to the classifier, as some of the features may overlap or convey 'redundant' and/or misleading information. In our work (chapters four and seven), the statistical measurements selected belong to different groups of statistics (i.e. location statistics, distribution statistics, etc.), so as to avoid the aforementioned issue of overlapping as discussed in chapter four.

2.3 The phase information

As mentioned in the introductory part of this chapter, the majority of the features employed for audio classification are time-domain or magnitude spectrum related. However, phase information maybe important in discriminating between acoustic classes as the results reported in chapter seven.

As will be discussed in chapter five (phase analysis and preliminary results), the phase function suffers from many discontinuities caused by computational artifacts and certain
properties that the signal conveys. These discontinuities affect any dimensionality reduction method (via a transform or via statistical measurements) applied to the phase spectrum in order to extract features that will be introduced to the classifier. Hence, these discontinuities have to be compensated/removed from the phase spectrum (chapters four, five and seven).

In the remaining part of this section, there will be an overview of the research papers that have been published in the area of phase analysis of 1D (one dimensional) signals. Until recently, the majority of the published work was focused on synthetic signals [88] however, in the last two years, a few papers investigate the phase content of speech signals, adopting an approach similar to our work (i.e. phase extraction from non-synthetic signals). Note that, the literature review provided in this section, is focused on the phase analysis of 1D signals only, as the phase analysis of 2D (two dimensional) signals involves apriori knowledge of certain characteristics of the system [45]. Hence, the published literature, which is related to the phase analysis of 1D signals, can be divided into three categories.

The first category includes the research that is based on the paper that addressed the issue of phase ‘unwrapping’ [41]. The compensation rule aims to restrict the phase transitions that are greater than $\pi$. However, as will be discussed in chapter five, the manner in which the Tribolet’s algorithm compensates for the ‘wrapping’ ambiguities is heuristic and thus, the compensation is not always appropriate. Based on [41], three other papers [28], [32] and [33] have been published. These papers provide an improved compensation rule compared to [41], however the approach is still heuristic and hence all the phase transitions are corrected in the same manner, irrespective of the source that causes them. Moreover, two recent publications, [34] and [40], are adopting frequency estimation techniques so as to evaluate the phase spectrum of the signal. However, these techniques are appropriate only for cases of deterministic signals.

The second category of phase analysis publications, discriminates between the two sources that cause the phase discontinuities. The first source, also addressed by [41], is
related to the use of the inverse tangent function whereas, the second source, is caused by the simultaneous approach to zero of the real and the imaginary components of the Fourier transform of the signal (i.e., ‘zeros’/‘poles’ located on the circumference of the unit circle). This latter type of discontinuity, which appears in the phase function, causes $\pi$ phase transitions that, can be compensated by the appropriate addition/subtraction of $\pi$ rather than multiples of $2\pi$ [41]. The publications that address this second source of phase discontinuity, which are related to the sign changes of the real and the imaginary components of the signal are [26], [27] and [31]. As it will be mentioned in chapter five, this latter type of discontinuity is called ‘intrinsic’, whereas the type of discontinuity addressed by [41] is called ‘extrinsic’. Similarly, [35] and [39] address the ‘intrinsic’ type of discontinuity and recommend the increase of the precision (number of decimal points), during the phase evaluation process, proportionally to the distance of a ‘pole’ / ‘zero’ from the circumference of the unit circle.

The third category of publications includes the research papers that cannot be organized in a single group. Thus, [36] and [37] describe the phase evaluation for the special case of deterministic signals where none of their ‘zeros’/‘poles’ are located on the circumference of the unit circle (i.e. lack of ‘intrinsic’ discontinuities). Moreover, [29] and [30], introduce the concept of phase ‘difference’ as a method in order to avoid the ‘wrapping’ ambiguities. Also, [38] proposes the phase reconstruction of deterministic signals via higher order spectral analysis. It is important to note that the published literature, related to all these three categories of phase analysis, includes examples mainly of deterministic signals. However, during the past two years, there are a few published papers that highlight the importance of phase information in speech processing [42], [44], [61], as well as the effect of noise in the phase spectrum of speech [43].
2.4 Summary of the theoretical overview

In conclusion, this chapter that forms the theoretical overview of the thesis, is divided into two parts. The first part provides a synopsis of the most popular feature extraction techniques for audio classification. The second part provides an overview of the published literature that is related to the phase analysis of one dimensional signals.

Hence, the first category of the features, analyzed in section 2.1, are extracted from the time domain (volume, ZCR, LPC), whereas the second category are extracted from the frequency domain (Pitch, Cepstral Coefficients, mel-cepstral coefficients). As mentioned, volume conveys information related to the amplitude of the signal and, it is not a useful feature for audio classification of recordings with similar shape of waveform in the time domain. Moreover, LPC, cepstral coefficients and MFCCs make certain assumptions that are suitable for speech related problems. Also, pitch is only appropriate for the discrimination between harmonic and non-harmonic classes of audio. The ZCR is an indication of the signal’s noise component and hence, it is not a useful feature for audio classification of recordings with similar noise content. However, the features that can be extracted from the magnitude spectrogram of a signal are useful for the application of ‘fine’ audio classification. Hence, in chapter seven (audio classification application), statistical features will be extracted from the time – frequency signal representations (including the Fourier magnitude spectrogram) so as to form the feature vectors that will be introduced to the classifier.

The second part of this chapter (section 2.3), provides an overview of the literature related to phase analysis of 1D signals. Only recently, researchers investigated the importance of phase in non-synthetic signals and have shown its significance in the area of speech processing. In chapters five, six (time localization of a signal application) and seven (audio classification application) features will be extracted from the phase spectrum in the appropriate manner, so as to be used for the time-delay measurements (chapters five and six) and the audio classification application (chapter seven).
CHAPTER 3
TRANSFORMS AND TIME-FREQUENCY DISTRIBUTIONS

Introduction

In this chapter, the Hartley transform and the time-frequency distributions used for this research will be introduced. The relationship between the Hartley and the Fourier spectrum will be established. The importance of the time-frequency signal representation will be highlighted, as well as a brief description of the magnitude spectrogram and of the Wigner-Ville distribution.

Hence, section 3.1 provides details related to the Hartley transform, such as its mathematical relationship with the Fourier transform as well as a short-time analysis of the Hartley phase spectrum ('whitened' Hartley spectrum). In section 3.2, a brief description of the homomorphic deconvolution process is provided. The Fourier and the Hartley phase cepstrums are used for the time localization of a signal application, which will be presented in chapters five and six. Sections 3.3, 3.4 and 3.5, are dedicated to the
description of the time-frequency signal representation, the quadratic time-frequency signal representation and the visualisation of a time-frequency representation, respectively. Hence, sections 3.3 and 3.4 are related to chapter seven, where statistical features are extracted from the time-frequency signal representations so as to form the feature vectors introduced to the classifier for the audio classification application. Finally, section 3.6 provides the summary of this chapter.

3.1 The Hartley transform

In this section, the relationship between the Hartley spectrum [67], [71], [72] [99] and the Fourier spectrum [66], [92] is established (subsection 3.1.1), the Discrete-time Hartley transform (DTHT) is presented (subsection 3.1.2) and finally, a short-time analysis of the Hartley phase spectrum is provided (subsection 3.1.3).

3.1.1 The Hartley/Fourier spectral relationship

In this subsection the relationship of the magnitude and phase spectra of the Fourier and Hartley transforms [91] will be established. From [73], the relationship between the two transforms can be expressed as:

\[ H(\omega) = \Re(S(\omega)) - \Im(S(\omega)) \]  \hspace{1cm} (3.1)

where \( H(\omega) \) is the Hartley spectrum, \( \Re(S(\omega)) \) is the real component and \( \Im(S(\omega)) \) is the imaginary component of the Fourier transform \( S(\omega) \), respectively.

The Hartley magnitude spectrum and the ‘whitened’ Hartley spectrum (Hartley phase spectrum) will now be defined based on equation (3.1). The definition of the Fourier magnitude and phase [74] are:

\[ M(\omega) = \sqrt{\Re^2(S(\omega)) + \Im^2(S(\omega))} \hspace{1cm} (3.2) \]

\[ \varphi(\omega) = \arctan\left(\frac{\Im(S(\omega))}{\Re(S(\omega))}\right) \hspace{1cm} (3.3) \]

Also,
Chapter 3. Transforms and time-frequency distributions

\[ S(\omega) = M(\omega)(\cos(\varphi(\omega)) - j \sin(\varphi(\omega))) \] (3.4)

Moreover, \( \Re(S(\omega)) = M(\omega)\cos(\varphi(\omega)) \) (3.5) and
\[ \Im(S(\omega)) = -M(\omega)\sin(\varphi(\omega)) \] (3.6)

thus from equation (3.1),
\[ H(\omega) = \Re(S(\omega)) - \Im(S(\omega)) \]
\[ = M(\omega)(\cos(\varphi(\omega)) + \sin(\varphi(\omega))) \] (3.7).

Also, \( H^*(\omega) = \Re(S(\omega)) + \Im(S(\omega)) \)
\[ = M(\omega)(\cos(\varphi(\omega)) - \sin(\varphi(\omega))) \] (3.8)
where * denotes complementation.

Concluding, the Hartley magnitude is defined as:
\[ N(\omega) = \sqrt{H(\omega)H^*(\omega)} = M(\omega)\sqrt{\cos^2(\varphi(\omega)) - \sin^2(\varphi(\omega))} = M(\omega)\sqrt{\cos(2\varphi(\omega))} \] (3.9)

Thus, the Hartley magnitude spectrum is a combination of the Fourier magnitude spectrum \( M(\omega) \) and the Fourier phase spectrum \( \varphi(\omega) \). Hence, in a single formula, both the Fourier magnitude and the Fourier phase content of the signal is encapsulated.

Then, the 'whitened' Hartley spectrum is defined as:
\[ Y(\omega) = \frac{H(\omega)}{M(\omega)} = \frac{M(\omega)(\cos(\varphi(\omega)) + \sin(\varphi(\omega)))}{M(\omega)} = \cos(\varphi(\omega)) + \sin(\varphi(\omega)) \] (3.10)

The last equation is a function of the Fourier phase \( \varphi(\omega) \) only and is called the 'scaled'/'whitened' Hartley spectrum [63], [64] or the Hartley phase spectrum. Equation (3.10) shows that the 'whitened' Hartley spectrum is equal to the addition of the cosine of the Fourier phase with the sine of the Fourier phase and thus, it conveys Fourier phase
information. From the definition of the 'whitened' Hartley spectrum, (equation (3.10)), it can be shown that it has upper and lower bounds of $\pm \sqrt{2}$ [63].

3.1.2 The Discrete Hartley transform

As with the DTFT [83], the discrete-time, continuous frequency, Hartley transform (DTHT) [75], is defined as:

$$X(f) = \sum_{n=0}^{N-1} x(n) (\cos(2\pi fn) + \sin(2\pi fn))$$  \hspace{1cm} (3.11)

where $x(n)$ is a finite length discrete signal, $n \in [0, N-1]$ and $f$ is continuous frequency. The Inverse Discrete-Time Hartley Transform (IDTHT) is defined as:

$$x(n) = \frac{1}{N} \int_{-\pi}^{\pi} X(f) (\cos(2\pi fn) + \sin(2\pi fn)) df$$  \hspace{1cm} (3.12)

Thus, a full discrete implementation of the DTHT is the Discrete-Hartley Transform (DHT) and can be defined as:

$$X(k) = \sum_{n=0}^{N-1} x(n) \left( \cos \frac{2\pi kn}{N} + \sin \frac{2\pi kn}{N} \right)$$  \hspace{1cm} (3.13)

for $k = 0, 1, ..., N - 1$

and the Inverse Discrete-Hartley Transform (IDHT) is defined as:

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) \left( \cos \frac{2\pi kn}{N} + \sin \frac{2\pi kn}{N} \right)$$  \hspace{1cm} (3.14)

for $n = 0, 1, ..., N - 1$

where $n$ and $k$ are the time and frequency indices, respectively.
3.1.3 Short-time Analysis and the Hartley phase spectrum

In the short-time analysis of the finite length discrete-time signals the effect of windowing (see also section 4.2 of chapter four) is critical. Assuming that the local time-axis is taken as zero at the centre of the windowed data, then the corresponding frequency spectrum can be represented by a simple convolution of the two spectra. Thus,

\[ H[s(t)] = H[s(t)w(t)] = H[s(t)] * H[w(t)] \]  \hspace{1cm} (3.15)

where \( \hat{s}(t) \) is the short-time frame of the signal \( s(t) \).

Similarly, for the Fourier case

\[ F[\hat{s}(t)] = F[s(t)w(t)] = F[s(t)] * F[w(t)]. \]

Assuming that the window, \( w(t) \), has even symmetry, i.e., consider the zero time-axis at the centre of the window, then

\[ F[w(t)] = H[w(t)] = W(\omega) \]

which is a real and even function of frequency.

For the case of the magnitude spectrum of the windowed signal,

\[ \hat{M}(\omega) = |F[s(t)] * W(\omega)| \]  \hspace{1cm} (3.17).

Moreover, \( F[s(t)] = \Re(S(\omega)) - j\Im(S(\omega)) \) and thus, equation (3.17) becomes

\[ \hat{M}(\omega) = \sqrt{\Re(S(\omega))^2 + \Im(S(\omega))^2} W(\omega) \]

Assuming that the cross-correlation between the window function and the time signal is zero, then the last equation becomes

\[ \hat{M}(\omega) = \sqrt{\Re^2(S(\omega)) + \Im^2(S(\omega))} W^2(\omega) \]

\[ = M(\omega) * W(\omega) \]  \hspace{1cm} (3.19)
For the case of the Hartley phase spectrum of the windowed signal, based on equation (3.10),

\[
\hat{Y}(\omega) = \frac{H[s(t)]}{M(\omega)}
\]  

(3.20)

Also, based on equations (3.15) and (3.19), the last equation becomes

\[
\hat{Y}(\omega) = \frac{H[s(t)]*W(\omega)}{M(\omega)*W(\omega)}
\]

\[
= \left( \frac{M(\omega)Y(\omega)}{M(\omega)*W(\omega)} \right) * W(\omega)
\]

(3.21)

Equation (3.21) presents the Hartley phase spectrum of the windowed signal. However, since this is the first time where the short-time analysis of the Hartley phase spectrum is presented, the issue of the different effect that windowing has to different spectra, will be emphasized. Hence, assume that there are two window spectral functions \( W^a(\omega) \) and \( W^b(\omega) \). Also, based on the original \( W(\omega) \) window function, assume that, \( 0 < a, b \leq 1 \) and \( a + b = 1 \). This assumption is considered because windowing affects the Fourier magnitude spectrum differently compared to the Hartley phase spectrum, depending on the correlation between the window and each of the spectra. Thus, the numerator of equation (3.21) becomes,

\[
(M(\omega)Y(\omega)) * W(\omega) = (M(\omega) * W^a(\omega)) * W^b(\omega)
\]

(3.22)

Note that the last approximation holds in the case where \( M(\omega) \) and \( Y(\omega) \) are uncorrelated \[62\] [106].

Consequently, (3.30) becomes,

\[
\hat{Y}(\omega) = \epsilon(\omega) \left[ Y(\omega) * W^b(\omega) \right]
\]

(3.23)

where
Chapter 3. Transforms and time-frequency distributions

\[ \varepsilon(\omega) = \frac{M(\omega) * W^n(\omega)}{M(\omega) * W(\omega)} \]  

(3.24)

The effect of \( \varepsilon(\omega) \) to the Hartley phase spectrum depends on the features of the Hartley magnitude spectrum, the features of the signal used and the features of the window employed [107].

### 3.2 The homomorphic deconvolution process

In order to present the phase content of a signal in a useful manner, it is necessary to choose a function that can encapsulate it. Finding the cepstrum of the phase function does this. This class of methods is mentioned in literature as a homomorphic deconvolution process [100]. The homomorphic process describes the invertible procedure in which a signal is transformed into another domain, employing an orthogonal transform \( \Xi \), and then after applying a non-linear process to that new domain, is transformed back, via \( \Xi^{-1} \), to the original domain, which is now called the cepstral domain.

The figure 3.1 summarizes the homomorphic deconvolution process.

![Figure 3.1: Summary of the homomorphic deconvolution process](image)

Hence, the Fourier phase cepstrum is defined as:
3.2 \( c_f(\tau) = \text{IDTFT}(\phi(\omega)) \) (3.25)
i.e. the inverse Fourier transform of the Fourier phase spectrum of the signal

while, the Harley phase cepstrum is defined as:

\[ c_H(\tau) = \text{IDTHT}(Y(\omega)) \] (3.26)
i.e. the inverse Hartley transform of the Hartley phase spectrum of the signal.

The Fourier and the Hartley phase cepstrums will be used for the time localization of a
signal application, which is presented in chapters five and six.

3.3 Time-frequency signal representation

Non-synthetic signals are characterized by non-stationarity, i.e., their spectra evolve in
time. Consequently, in the case where a transform is applied to the whole length of the
time signal, this results in many local features being lost or smoothed due the averaging
process. Thus, the time-frequency signal representation aims to investigate the properties
of the signal in the frequency domain without suppressing the signal’s local
characteristics due to the application of the transform to the whole length of the signal.

Hence, the discrete-time signal is divided into frames of equal length each, and then the
transform is applied to each frame independently. The general algorithm for
implementing the time-frequency representation of a signal can be summarized as
follows. Initially, a time domain window is localized to the centre of the time domain
data frame that will be spectrally analysed. The Fourier transform of the multiplication of
the window function with the time domain data frame, results in the localized frequency
representation of the signal. Subsequently, the same window is slid across each frame of
the time signal. Consequently, a matrix is constructed from the discrete form of this
process. The first row of the matrix corresponds to the spectrum of the first time frame
while, the last row of the matrix, to the spectrum of the last time frame. Hence, the time-
frequency signal representation presents the evolution of the frequency content of the
signal with respect to time [90]. Note that, important factors that affect the time-
frequency signal analysis are the window size and the shape of the window. The time –
frequency signal representations (sections 3.3 and 3.4) are employed in chapter seven
Chapter 3. Transforms and time-frequency distributions

(audio classification application) where statistical features are extracted from them so as to form the feature vectors introduced to the classifier.

Perhaps the most common form of time-frequency representation is derived from the Fourier transform and is usually referred to as the Fourier magnitude spectrogram. The short time Fourier transform (STFT) is defined as:

\[ S(f, \tau) = \int_{-\infty}^{\infty} s(t)w(t-\tau)e^{-j2\pi ft} \, dt \quad (3.27) \]

where \( s(t) \) is the time-domain signal, \( S(f, \tau) \) is the short-time estimate of the Fourier transform and \( \tau \) represents the position of the window on the time axis. Note that the last formula is a function of both frequency and time, where the time dependency is related to the window location.

Moreover,

\[ S(f, \tau) = F[s(t)]*F[w(t-\tau)] \]

Substituting,
\[ F[j(f)] = X(A) \quad \text{and} \quad F[w(t-\tau)] = F[W(\lambda)]e^{-j2\pi \lambda \tau} = W(\lambda)e^{-j2\pi \lambda \tau}, \]
where \( \lambda \) represents an intermediate frequency variable, gives us:

\[ S(f, \tau) = X(\lambda)^* (W(\lambda)e^{-j2\pi \lambda \tau}) \]
\[ = \int_{A} X(\lambda)W^*(\lambda-f)e^{j2\pi (\lambda-f) \tau} d\lambda \]
\[ = e^{-j2\pi \tau} \int_{A} X(\lambda)W^*(\lambda-f)e^{j2\pi \lambda \tau} d\lambda \quad (3.28) \]

where, \((.)^*\) denotes conjugation.

The last equation presents the frequency domain version of the time-frequency function. Analysing equation (3.28) and excluding the \( e^{-j2\pi \tau} \) term, one could interpret the time-
The frequency function of equation (3.28) as the ISTFT (inverse short-time Fourier transform) of \( X(\lambda) \). The equivalent expression for the time domain version of the time-frequency function is:

\[
S(f,\tau) = \int_{-\infty}^{\infty} s(t)w^*(t-\tau)e^{-j2\pi ft}dt 
\tag{3.29}
\]

The relationship between equations (3.28) and (3.29) addresses the issue of resolution in the time and frequency domains. Hence, the narrower the time domain window the higher the resolution in the time domain. Analogously, the narrower the frequency domain window, the higher the resolution provided in the frequency domain. However, the resolutions of the time and frequency domains are inversely proportional. This is an example of the well-known Heisenberg-Gabor principle of uncertainty when applied to the area of signal processing. The proof of that theorem is provided in section 4.2 of chapter four.

### 3.4 Quadratic time-frequency signal representation

The complex time-frequency representation of a signal, based on the Fourier Transform, belongs to the category of the Linear Time-Frequency Representations (LTFRs). Analogously, the Fourier magnitude spectrogram belongs to the category of the Quadratic Time-Frequency Representations (QTFRs), taking their name after the presence of the squared terms.

An important property of a QTFR is the property of superposition, which is proved as follows. From the property of linearity, the summation of two time signals \( x(t) \) and \( y(t) \) is expressed in the frequency domain as:

\[
S(f) = X(f) + Y(f) 
\tag{3.30}
\]
where $X(f)$ and $Y(f)$ is the Fourier transform of the time signals $x(t)$ and $y(t)$, respectively. Both signals $X(f)$ and $Y(f)$ have a real and imaginary component and consequently, the power spectrum of the summation of these two signals, $P(f)$, can be expressed as:

$$P(f) = (\Re(S(f)) + j\Im(S(f)))(\Re(S(f)) - j\Im(S(f)))$$

$$= \Re^2(S(f)) + \Im^2(S(f)) = (\Re(X(f)) + \Re(Y(f)))^2 + (\Im(X(f)) + \Im(Y(f)))^2$$

$$= (\Re^2(X(f)) + \Im^2(X(f)) + \Re^2(Y(f)) + \Im^2(Y(f))) + 2\Re(X(f))\Re(Y(f)) + 2\Im(X(f))\Im(Y(f))$$

$$= M_x(f) + M_y(f) + 2\Re(M_{xy}(f)) \quad (3.31)$$

where $M_x(f)$, $M_y(f)$ denotes the magnitude spectrogram of $x(t)$ and $y(t)$, respectively and $\Re(M_{xy}(f))$ the real part of the 'cross-magnitude spectrogram'. The last relationship shows that the power spectrum resulted from the addition of two signals equals the summation of the power spectra of these signals with their cross-spectral term. The last relationship is known as the quadratic superposition principle. Note that, the existence of the cross-spectral term affects the separation of these two signals because more components appear in the time-frequency plane.

A popular QTFR is the Wigner-Ville distribution [69]. The advantages of the Wigner-Ville distribution, compared to the Fourier magnitude spectrogram, is its higher resolution and the fact that it does not represent only the magnitude content of the signal. The Wigner-Ville distribution is defined as:

$$W_D(t,f) = \int x\left(t + \frac{\tau}{2}\right)x^*\left(t - \frac{\tau}{2}\right)e^{-j2\pi\tau}d\tau \quad (3.32)$$

and

$$W_D(t,f) = \int x\left(f + \frac{\nu}{2}\right)x^*\left(f - \frac{\nu}{2}\right)e^{j2\pi\nu}d\nu \quad (3.33)$$
Two of the properties that makes the Wigner-Ville distribution superior compared to other QTFRs (e.g. Fourier magnitude spectrogram), are its marginal properties, expressed as:

\[ \int T(f, \tau) df = |s(t)|^2 \quad (3.34) \text{ (instantaneous signal power)} \]

\[ \int T(f, \tau) d\tau = |S(f)|^2 \quad (3.35) \text{ (power spectrum of the signal)} \]

where \( T(f, \tau) \) represent the Wigner-Ville distribution of a signal \( s(t) \).

Note that, the quadratic superposition principle (equation (3.31)), has the following form when applied to the Wigner-Ville case:

\[ P(f) = W_{Dx}(t, f) + W_{Dy}(t, f) + 2\Re(W_{Dx,y}(t, f)) \quad (3.36) \]

where, \( W_{Dx}(t, f) \), \( W_{Dy}(t, f) \) denotes the Wigner-Ville distribution of \( x(t) \) and \( y(t) \), respectively and \( \Re(W_{x,y}(t, f)) \) the real part of the 'cross-Wigner-Ville distribution' .

Note that, the cross interference term of the Fourier magnitude spectrogram (equation (3.31)) is different from the Wigner-Ville (equation (3.36)) one. Hence, for the Fourier magnitude spectrogram case, the interference terms are limited to the areas of the time-frequency plane that the two power spectra, \( M_x(f) \) and \( M_y(f) \), overlap. Consequently, for the Fourier magnitude spectrogram, in the case where the two signals are well separated, the cross-term is negligible. Although the last remark appears to be an advantage of the Fourier magnitude spectrogram, however this occurs due to its poor resolution.

As a reminder, the Wigner-Ville is a quadratic function and hence its sampling has to be done in a careful manner. Thus, from equation (3.32):
$$W_D(t, f) = \int_{-\infty}^{\infty} x\left(t + \frac{\tau}{2}\right)x^*\left(t - \frac{\tau}{2}\right)e^{-j2\pi f \tau} d\tau = 2 \int_{-\infty}^{\infty} x(t + \tau)x^*(t - \tau)e^{-j4\pi f \tau} d\tau \quad (3.37)$$

If $x$ is sampled with a period $T_e$, then $x(n) = x(nT_e)$ and consequently, the Wigner-Ville is evaluated at the sampling points $nT_e$. Hence, the discrete-time version of the Wigner-Ville can be expressed as:

$$W_D(n, f) = 2T_e \sum_k x[n + k]x^*[n - k]e^{-j4\pi f k} \quad (3.38)$$

Note that, in practice, the signals do not tend to infinity and consequently, they are windowed. The windowing of the Wigner-Ville distribution (pseudo-Wigner-Ville distribution) has smoothing effects and consequently the aforementioned marginal properties are lost. However, the resolution of the pseudo-Wigner-Ville distribution is still superior when compared to the resolution of the Fourier magnitude spectrogram.

A more detailed analysis of the Wigner-Ville distribution and its properties, as well as details about the Cohen’s class time-frequency distributions, in which the Wigner-Ville distribution belongs, can be found in [69] and [70].

### 3.5 Visualisation of a time-frequency representation

In this section an example of the Fourier magnitude spectrogram of an audio recording is provided. The Fourier magnitude spectrogram, $S(f, \tau)$, can be visualized as a continuous, real-valued function mapped on the time-frequency plane. The Fourier magnitude spectrogram provides a representation of the evolution of the signal’s magnitude across time.

The Fourier magnitude spectrogram of a referee whistle blow (football match) is presented in figure 3.2. One can observe that the dominant frequency appears at around 0.18 of the frequency scale. Note that the magnitude intensity is proportional to the colour intensity.
Figure 3.2: Fourier magnitude spectrogram of a referee whistle

A line drawn perpendicularly to the x-axis of the time-frequency distribution expresses an estimation of the magnitude content of the signal at a specific time whereas, a horizontal line, drawn perpendicularly to the y-axis, provides an estimation of the evolution of the magnitude content of the signal at a certain frequency.

### 3.6 Summary

In this chapter, the Hartley transform was briefly described. In subsection 3.1.3, a short-time analysis of the Hartley phase spectrum (‘whitened’ Hartley spectrum) was provided. In section 3.2, the Fourier and the Hartley phase cepstrums, which will be used for the time localization of a signal application (chapters five and six), were defined. The advantage of the time-frequency signal representation, in terms of preservation of the signal’s features, was highlighted (section 3.3). In section 3.4, some of the advantages of the Wigner-Ville signal representation, were mentioned. The Wigner-Ville distribution provides higher resolution compared to the Fourier magnitude spectrogram and also
encapsulates the signal's characteristics to a single plot, unlike the Fourier magnitude that encapsulates only the magnitude content of the signal. Hence, as will be discussed in chapter seven, statistical measurements extracted from the time-frequency signal representations will be used in order to form the feature vectors, which will be introduced to the classifier for the audio classification application (chapter seven).
CHAPTER 4
CLASSIFICATION AND OTHER PRACTICAL CONSIDERATIONS

Introduction

In developing the audio classification software (chapter seven), there were a number of implementation issues that needed to be addressed. In this chapter, these issues have been brought together. Section 4.1 discusses the issue of the signal sampling rate. The phenomenon of windowing is described in section 4.2, while in section 4.3, the issue of 'root' finding of high-order polynomials is analyzed. Section 4.4 discusses the statistical measurements used while, section 4.5 examines the effect of the use of 'difference' when extracting statistical features from the phase related spectrograms. Finally, section 4.6 is dedicated to the distance-metric classifiers.

This chapter is related to chapters five and seven in the following manner. Section 4.2 describes the effect of windowing in the case where the audio recordings (chapter seven) of the database used are divided into frames so as to structure the time – frequency
signal representations from which the statistical measurements (section 4.4) will form the feature vectors introduced to the distance metric classifier (section 4.6). Moreover section 4.3 is related to the evaluation of the phase spectrum of a signal via the z-transform that will be discussed analytically in chapters five and seven.

4.1 The sampling process

In this first section, the issue of the sampling rate is addressed. In theory, there is not a rule that describes the appropriate sampling rate of a non-periodic signal of finite length. In chapter seven though, the experimental results show that the classification scores are influenced by the sampling rate. Thus, the issues of undersampling and oversampling are discussed in the following subsections.

4.1.1 The sampling process and aliasing

In practise, a continuous signal can be expressed in its discrete form, if multiplied by equally spaced dirac delta functions, as shown below,

\[ s(n) = s(t) \sum_{n=-\infty}^{\infty} \delta(t-nT) \quad (4.1) \]

where, \( s(n) \) and \( s(t) \) are the discrete and continuous version of the signal respectively, \( \delta(t) \) represents the dirac delta function and \( T \) the sampling period. The dirac delta function, \( \delta(t) \), is defined as the limit in which the rectangular pulse of unit area tends, as the width of the pulse tends to zero, maintaining the area to unity [111]. Therefore, the Fourier transform of \( s(n) \) is:

\[
F[s(n)] = \int s(t) \sum_{n=-\infty}^{\infty} \delta(t-nT)e^{-j2\pi ft} dt \quad \text{or}
\]
\[ F[s(n)] = F[s(t)] \ast \sum_{-\infty}^{\infty} \delta(t - nT)e^{-j2\pi ft} \]  \hspace{1cm} (4.2)

where \( \ast \) denotes convolution. Focusing on the second part of equation (4.2),

\[ \sum_{-\infty}^{\infty} \delta(t - nT)e^{-j2\pi ft} dt = \sum_{n=-\infty}^{\infty} [\delta(t - nT)e^{-j2\pi ft} dt] = \sum_{n=-\infty}^{\infty} e^{-j2\pi ft} \]  \hspace{1cm} (4.3)

Also,

\[ \sum_{n=-\infty}^{\infty} e^{-j2\pi ft} = \sum_{n=-\infty}^{\infty} (\cos(2\pi fnT) - j \sin(2\pi fnT)) \]  \hspace{1cm} (4.4)

In the case where \(-\infty < n < \infty\) the summation of the sine terms of the last equation tends to zero, whereas the summation of the cosine terms tends to a set of periodic dirac delta functions in the frequency domain, i.e.

\[ \sum_{n=-\infty}^{\infty} (\cos(2\pi fnT)) = \sum_{k=-\infty}^{\infty} \delta(f - kf_s) \]  \hspace{1cm} (4.5)

where \( f_s = \frac{1}{T} \) represents the sampling frequency and \( k \) is an integer incrementing in the frequency domain.

Subsequently, equation (4.2) becomes,

\[ F[s(n)] = S(f) \ast \sum_{k=-\infty}^{\infty} \delta(f - kf_s) = \sum_{k=-\infty}^{\infty} [S(f) \ast \delta(f - kf_s)] \]  \hspace{1cm} (4.6)

Convolution of a signal with the dirac delta function does not have any effect to the signal. Thus, the convolution of a signal with frequency shifted dirac delta functions, results in a replication of the spectrum at a regular distance of \( f_s \) across the frequency axis. Note that, in the replicated spectrum each dirac delta function is frequency \( f_s \) away from each neighbouring dirac delta function. If the spectrum of the initial signal \( s(t) \) has
Chapter 4. Classification and other practical considerations

infinite bandwidth, then the replicated spectra overlap each other causing the aliasing effect. In order to avoid aliasing, the continuous signal has to be band-limited so as to remove all frequencies above or below \( f_s/2 \) and \(- f_s/2\), respectively. The last statement is widely known as the Nyquist limit which says that a continuous signal can be exactly recovered from its sampled counterpart, provided the bandwidth of the signal is less than half the sampling frequency. The last statement also highlights the importance of preprocessing the continuous signal before the sampling process takes place [90].

An important practical issue that arises at this stage is what happens in the case where the sampling rate is irregularly spaced. Thus, equation (4.1) now becomes,

\[
s(n) = s(t) \sum_{n=-\infty}^{\infty} \delta(t - nT \pm \varepsilon) \quad (4.7)
\]

where, \( \varepsilon \) represents a small and random variation in the position of each dirac delta function. Consequently, although the result is again a set of replicated spectra, the dirac delta functions are not uniformly separated any more; i.e.:

\[
f_s \pm \lambda = \frac{1}{T \pm \varepsilon}
\]

However, although the dirac delta functions are randomly separated across the spectra, the signal can still be accurately represented. Further restriction of the signal’s bandwidth, proportional to the randomness of \( \varepsilon \), will reduce the possibility of aliasing. The assurance of the lack of aliasing effects, is one of the reasons that the signal bandwidth has to be restricted, even further, compared to the Nyquist criterion.

4.1.2 The sampling process and the DFT

As mentioned before, in order to avoid aliasing in the frequency domain, the signal has to be bandwidth limited. Analogously, in order to avoid aliasing in the time domain, the signal has to be time limited. Hence, it is important to specify the number of the time
domain points required so as to avoid aliasing.

Due to reciprocity between the forward and the inverse Fourier transform, it is stated that for \( f_s N \) (\( N \): number of samples) samples taken from the discrete time domain of a signal, \( N \) samples from the frequency domain are required so as to represent the signal accurately. Therefore, in the frequency domain, there will be \( f_s N \) real spectral samples and \( f_s N \) imaginary spectral samples i.e. totally \( 2f_s N \). However, the real and the imaginary components of the complex spectrum of a real signal are characterized by even and odd symmetry about the amplitude axis, respectively. So, only \( f_s \frac{N}{2} \) frequency samples are required by each component (i.e., total number of frequency samples equals to, \( f_s N \)) so as the spectrum to be fully represented [87].

Taking only a limited number of time samples, in order to construct the spectrum, affects its resolution, therefore the new spectrum is just an estimation of the ‘original’ one. However, even if the sampling rates in both the time and frequency domains are ‘appropriate’, the observations about the signal characteristics are usually derived from either the magnitude or the phase spectra. The evaluation of the Fourier magnitude (equation (3.2), chapter three) and the Fourier phase (equation (3.3), chapter three) spectra though, involve the use of non-linear functions. Observations indicate that even in the case where the complex spectrum is sampled correctly, the magnitude and phase spectra are oversampled and undersampled, respectively. Hence, there is no analytical method that guarantees the correct sampling of the magnitude or the phase spectra. In most applications, the effect of oversampling is less critical compared to the effect of undersampling.

4.2 Windowing

As will be discussed in chapter seven, each audio recording of the database has to be divided into frames (windowing) so as to form the time – frequency distributions. The
effect of windowing, in signal processing, is similar to the observation of an infinitely long signal through a ‘window’.

The window is defined as a time signal with finite positive amplitude spanned over a certain time period whereas elsewhere equals zero. If a certain frame of a time data signal, \( x(t) \), is windowed by \( w(t) \), then the new signal \( x'(t) \), in the time domain, is formed as:

\[
x'(t) = x(t)w(t) \quad (4.8),
\]

The Fourier transform of that expression is,

\[
X'(f) = S(f)*W(f) \quad (4.9)
\]

From the last relationship, it can be seen that the spectrum of the windowed signal \( x'(t) \) is equal to the convolution of the ‘actual’ spectrum of the signal with the spectrum of the window (subsection 3.1.3, chapter three). Thus, the spectrum of the windowed data strongly depends upon the spectral properties of the windowing function. For example, the Fourier transform of a rectangular window function results in a sinc function (i.e., \( \frac{\sin(f(x))}{f(x)} \)). The sinc function has a narrow dominant central lobe and significant side lobes that affect the neighbouring spectral frames, as they diffuse their spectral leakage across the power spectrum. Hence, fine information of the neighbouring spectral frames will be suppressed resulting in the degradation of the information content of the spectrum. The effect of windowing becomes less apparent by choosing the appropriate window shape [62].

Moreover, another crucial point in the operation of windowing is its size in the time and frequency domain. The narrower the time domain window, the higher the resolution provided in the time domain. Similarly, the narrower the frequency domain window, the higher the resolution provided in the frequency domain. However, the resolutions of the time and frequency domains are inversely proportional [90]. The following example
shows the effect of this inverse proportionality for the case of a rectangular pulse.

The Fourier transform of a rectangular pulse of width equal to $2\delta\tau$ is expressed as:

$$W(f) = \frac{\sin(2\pi f \delta\tau)}{\pi f} \quad (4.10)$$

Equation (4.10) forms the sinc function, where its first zeros occur when

$$\sin(2\pi f \delta\tau) = 0, \quad 2\pi f \delta\tau = \pm\pi$$

giving:

$$f \delta\tau = \pm\frac{1}{2} \quad (4.11)$$

However, note that, $\delta f = 2f$, where $\delta f$ is the distance between the first two zeros, on either side of the main lobe, of the sinc function in the frequency spectrum. Hence, equation (4.11) becomes

$$\delta f \delta\tau = \pm 1 \quad (4.12)$$

The last expression shows that the window widths in the time and frequency domain are inversely proportional.

### 4.3 Evaluation of polynomial ‘roots’

In chapters five and seven, which are dedicated to phase analysis and to the experimental results, the issue of ‘root’ estimation of high-order polynomials arises. Specifically, in the phase spectrum evaluation via the z-transform (chapters five and seven), the signal is divided into frames of 256 samples and then each frame is treated as a polynomial of order 255. This polynomial consists of 256 coefficients, i.e., the maximum power of the polynomial equals 255. The ‘roots’ of this polynomial are evaluated and mapped onto the z-plane, in order to estimate the phase spectrum of that frame of data [66].

It is relatively easy, by appropriate manipulation of the general quadratic formula, to calculate the ‘roots’ of polynomials of degree four or less. However, there is not a
standard formula to calculate the 'roots' of a higher than fifth degree polynomial. In the case that a polynomial cannot be easily factorised there are numerical methods that calculate its 'roots'. A summary of these methods with their advantages and disadvantages can be found in [77]. Perhaps the most popular of the numerical methods are based on the Newton’s approach. However, they have the disadvantage of not always converging towards a 'root' or sometimes converge very slowly. Moreover, the same algorithm has to be repeatedly applied so as to detect all the 'roots' of the polynomial. An alternative method, based on linear algebra, calculates the 'roots' of a polynomial by estimating the eigenvalues of the companion matrix formed by the polynomial’s coefficients.

Hence, the aim is to find the 'roots' of the following polynomial of degree \( n \) that is of the form:

\[
a_0 + a_1 t + \ldots + a_{n-1} t^{n-1} + a_n t^n = 0 \quad (4.13)
\]

where \( a_0, a_1, \ldots, a_{n-1} \) and \( a_n \) are real coefficients and \( a_n \neq 0 \) or

\[
c_0 + c_1 t + \ldots + c_{n-1} t^{n-1} + t^n = 0 \quad (4.14)
\]

The QR (eigenvalue) method is selected as the 'roots' finding technique for the phase estimation via the z-transform (chapters five and seven). Further details regarding the QR algorithm can be found in [85] and in appendix A1.

### 4.4 Statistical measurements

This section discusses the statistical measurements extracted from the time – frequency signal representations of each audio recording, in order to form the feature vector(s) used for the audio classification application (chapter seven). Each of the time – frequency signal representations has to reduce its dimensionality in order to be introduced to the
classifier. Generally speaking, the effective dimensionality reduction is achieved based on either of the following two methods. The first method involves the selection of a certain number of coefficients derived from a transform with compression qualities (e.g. Discrete cosine transform) while the second, involves the extraction of a certain number of statistical measurements [13] [102]. As it will be revealed in chapters five and seven, the phase spectra (and consequently the phase related time-frequency signal representations) of the non-synthetic signals are characterized by rapid changes that appear as discontinuities. Consequently, it is not possible to apply any transform with compression qualities, because integration is involved which requires continuity. Therefore, for this research, the dimensionality reduction is achieved via the evaluation of statistical measurements.

The statistical features are calculated from each of the time-frequency signal representations. Each time-frequency representation forms a matrix. Hence, the statistical features are calculated from the distributions formed from the values of each of these matrices.

The descriptive statistics, are statistical measurements that extract features related to the distribution formed and are divided into five categories [80]. These categories are:

a) Location statistics: mean (1st order moment), median, geometric mean, mode, midrange, sum etc. They form the most popular category of statistics and describe the central tendency of the data under analysis [103].

b) Scale statistics or measurements of dispersion: variance (2nd order moment), standard deviation, range etc.

c) Shape and distribution statistics: skewness (3rd order moment, degree of asymmetry), kurtosis (4th order moment, degree of flatness) etc.

d) Quantiles: maximum, minimum, lower quartile, higher quartile, inter-quartile range etc. [101].

e) Counts: sample size, number missing, number of unique values etc.
Statistical measurements that belong to the same category convey similar information and therefore, certain measurements were chosen from each category. The statistical measurements selected to form the feature vectors for the audio classification application (chapter seven) are justified in the following subsections.

4.4.1 Location statistics

Perhaps the most popular location descriptor, the mean value, is not robust due to its susceptibility to outliers. Similarly, the sample mean, the geometric mean and the harmonic mean are not chosen for the same reason. The trimmed mean is also inappropriate, as the percentage of the distribution that has to be excluded, depends on the characteristics of each class of audio recordings. Hence, different classes would require different percentages of exclusion. Two other location statistics are the midrange and the sum. The midrange is defined as the average of the smallest and largest value of the sample data set. The sample sum is the sum of the values of all the data samples, i.e. same as the mean but multiplied by the sample size. Both descriptors, especially the first one, are heavily influenced by the outliers and consequently, they are not appropriate for distributions with unknown behaviour, as their values could be misleading and hence, confuse the classifier. The sample mode, another location statistical value, evaluates the data that occurs most frequently in the sample data set. The sample mode is less sensitive to the skewness of a distribution compared to either the sample mean or the sample median. However, the sample mode is more susceptible to the variation of the data sample when compared to both the sample mean and the sample median [103]. In the case of real signals, the variation of the data sample is an important factor that has to be taken into consideration. The median is a more robust statistical measurement compared to the arithmetic mean as it is calculated based on one or two values only (depending on the even or odd number of the data samples) [102]. This is the reason why median was chosen as one of the statistical measurements that formed the feature vector.
4.4.2 Scale statistics

The scale statistics provide information about the dispersion of the distribution i.e., the amount of scattering. The two most popular scale statistics are the sample variance together with its close relative the sample standard deviation that is the square root of the sample variance. Data outliers and skewness of the sample influence the sample variance and the sample standard deviation. However, the sampled variance is an unbiased statistical estimate of the distribution. The deviations of the mean absolute deviation (m.a.d.), although still influenced by extreme values, are not squared and consequently, outliers do not affect the result so heavily [102]. Other scale statistics are the sample error of the mean and the sample coefficient of variation. Both of them are characterized by lack of robustness, as their estimation is based upon the whole data set. Moreover, they convey information similar to the variance and standard deviation. Summarizing, from the scale statistics the variance, and the mean absolute deviation were chosen for the feature vector introduced to the classifier.

4.4.3 Shape and distribution statistics

From the category of shape and distribution statistics, the skewnesss and the kurtosis were chosen in order to form the feature vector introduced to the classifier. The shape statistics evaluate the difference, in terms of shape, between the data distribution under investigation and the normal distribution that has the same mean and variance to the one under investigation. The sample skewness and sample kurtosis are not robust statistical descriptors, as they are making use of all the data set, however they encapsulate distinctive features of the distribution efficiently [80].

4.4.4 Quantile statistics

Perhaps, the most popular quantile statistical descriptors are the minimum and maximum value of a distribution that are the $0^{th}$ and $100^{th}$ quantile, respectively. Moreover, the median, which also belongs to the location statistical descriptors, is the $50^{th}$ percentile.
Two other quantile statistical descriptors are the sample range and the inter-quartile range (i.q.r.). The range can be affected by outliers and is always used together with other statistical descriptors. The inter-quartile range is very robust to outliers compared to most of the statistical descriptors [101]. The range and the inter-quartile range (i.q.r.) were two of the statistical measurements chosen to form the feature vector.

**4.4.5 Count statistics**

The last category of the statistical descriptors is the counts. The sample size is the number of samples that forms the data set. The numbers missing are the values that are missing from the data set. Finally, the number of unique values is the number of values that are different compared to the rest of the data set. Due to the lack of robustness, none of the count statistical descriptors were chosen to contribute to the feature vector introduced to the classifier.

**4.4.6 Statistical feature vector**

Summarizing, in subsections 4.4.1 to 4.4.5, the statistical measurements were categorized and the most appropriate of them were identified for the audio classification application (chapter seven). Hence, in this subsection, the eight statistical measurements employed, to form the feature vector introduced to the classifier for the audio classification application (chapter seven), are defined. The reasons that these eight particular measurements are selected instead of others are described in the corresponding subsections (i.e. subsections 4.4.1 to 4.4.5).

i) Variance: is a measurement of the spread of a distribution and in particular, regarding the audio classification application (chapter seven), an indication of the dynamic range of the spectrum. It is defined as the mean of the squares of the differences between the data samples and the data mean value. Note that, instead of dividing the sum of the squares by the sample size, the sum of squares is divided by the sample size minus one (unbiased estimate) i.e.
\[ \text{variance} = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2 \]  

(4.15)

where, \( \bar{x} \) represents the sample mean and \( N \) the size of the data set.

\( \text{ii) Skewness:} \) is a measurement of asymmetry of the samples around the mean value and regarding the audio classification application (chapter seven), an indication of the shape of the spectrum. It is defined as the mean of the third power of the differences between the data samples and the data mean value.

\[ \text{skewness} = \frac{E(x - \bar{x})^3}{\sigma^3} \]  

(4.16)

where, \( \sigma \) represents the sample standard deviation.

\( \text{iii) Kurtosis:} \) is a measurement of the outlier proneness of a distribution and an indication of the shape of the spectrum for the audio classification application (chapter seven). It is defined as the mean of the fourth power of the differences between the data samples and the data mean value.

\[ \text{kurtosis} = \frac{E(x - \bar{x})^4}{\sigma^4} \]  

(4.17)

where, \( \sigma \) again, represents the sample standard deviation.

\( \text{iv) Inter-quartile range (i.q.r.):} \) is a measurement of dispersion and an indication of the dynamic range of the spectrum for the audio classification application (chapter seven). In most cases, the 50\% of the data located in the middle is the most important. The inter-quartile range is the difference between the 75\textsuperscript{th} and 25\textsuperscript{th} data percentile.

\[ \text{i.q.r.} = Q_{75} - Q_{25} \]  

(4.18)

\( \text{v) Median:} \) is a measurement of location as well as a quantile statistical measurement. In particular, for the audio classification application (chapter seven), median is an indicator of the central level of the spectrum. It is defined as the central value of the data set. Thus,
it divides the data set in half, fifty percent of the measurements being over it and fifty percent below it.

vi) Mean absolute deviation is a measurement of dispersion and particularly, regarding the audio classification application (chapter seven), an indication of the dynamic range of the spectrum. It evaluates the average of the absolute differences between the data samples and the mean of the data samples.

\[ m.a.d. = E \left| x_i - \bar{x} \right| \]  

(4.19)

where \( \bar{x} \) represents sample mean.

vii) Range: is the difference between the maximum and minimum values and regarding the audio classification application (chapter seven) an indication of the dynamic range of the spectrum.

viii) Entropy: expresses the information-related feature of the signal. For this work, the 'log energy' entropy is employed. Note that, entropy is an additive function

i.e.

\[ E(s) = \sum_i E(s_i) . \]

Thus, the 'log energy' entropy, is defined as:

\[ E(s_i) = \log(s_i^2) \]  

(4.20)

\[ E(s) = \sum_i \log(s_i^2) \]  

(4.21).

where \( s \) represents the signal and \( s_i \) the coefficients of \( s \).

Although entropy does not belong to any of the aforementioned categories of statistics (subsections 4.4.1 to 4.4.5), it was selected as one of the statistical measurements of the feature vector introduced to the classifier. Entropy has different physical meanings depending on the scientific area. In thermodynamics, it represents the amount of disorder in a system whereas, in communication theory represents the amount of uncertainty that a
signal conveys. Generally speaking, low entropy is observed when the amount of disorder or randomness is low [81], [84]. For the audio classification application (chapter seven), entropy conveys information related to the structural order of the spectrum.

### 4.5 Differential of the phase related spectra

The experimental results of the audio classification application (chapter seven) show that the classification rates reached in the case where the statistical features are calculated from the phase 'difference' spectrograms are higher when compared to the case where the classification rates are calculated from the phase spectrograms (i.e. without the use of 'difference'). Hence, in this section, initially the first differential (continuous domain) of the Fourier and the Hartley phase spectra and then, the first 'difference' (discrete domain) of these spectra, will be presented.

The first differential of the Fourier phase spectrum is:

\[
\frac{d(\phi(\omega))}{d\omega} = \frac{d \left( \arctan \left( \frac{\Im(S(\omega))}{\Re(S(\omega))} \right) \right)}{d\omega} = \frac{d\left(\Im(S(\omega))\right) \Re(S(\omega)) - \Im(S(\omega)) d\left(\Re(S(\omega))\right)}{\Re^2(S(\omega)) + \Im^2(S(\omega))} \tag{4.22}
\]

while, the first differential of the Hartley phase spectrum is:

\[
\frac{d(Y(\omega))}{d\omega} = \frac{d(\cos(\phi(\omega)) + \sin(\phi(\omega)))}{d\omega} = \frac{d(\phi(\omega))}{d\omega} \left(\cos(\phi(\omega)) - \sin(\phi(\omega))\right) \tag{4.23}
\]

In the discrete domain, the 'difference' of the Fourier phase spectrum is calculated by subtracting the first Fourier phase point from the second Fourier phase point, etc. along the discrete frequency axis. As with the 'difference' of the Fourier phase spectrum, the 'difference' of the Hartley phase spectrum is calculated by subtracting the first Hartley
phase point from the second Hartley phase point, etc. along the discrete frequency axis.

The classification rate is higher in the case where the 'difference' is used because the 'difference' emphasizes the information related to the phase spectral transitions of the phase spectrum (i.e., the more gradual the transition, the smaller the derivative amplitude and vice versa). Details related to the increase in the classification rate, obtained from the phase spectrograms, in the case where the 'difference' is used, will be presented in subsection 7.4.1 of chapter seven.

4.6 Distance metric classifiers

In this section, the choice of the classifier employed for the decision-making part of the pattern recognition process (chapter seven), is justified. There are many classification techniques in statistical pattern recognition; some of them are less complex, such as Template matching whereas, some others are more complicated such as the Support Vector Classifier [79], [82]. This work is focused on the feature extraction part of the pattern recognition process and hence, one of the least complicated classification techniques, the nearest mean, is used. The concept of the nearest mean classification technique, is based on the distance between each of the test vectors and the target vector. A test vector is classified to the class that corresponds to the target vector that is located closest.

There are many ways to compute the distance between two points. Five of the most popular metric distances are: the Euclidean, the standardized Euclidean, the City block, the Minkowski and the Mahalanobis [78] [79]. Note that, in most classification problems the target/test feature vectors cannot be visualized because they have more than three dimensions, however the principles of distance evaluation are the same.

Assume, an $M$-by-$N$ matrix where each of the rows ($m_i$) represents a different point in space and each of the columns ($n_j$) its coordinates. Generally, the metric distance, $d$, between two of the rows (e.g. $m_k$ to $m_l$) is defined as:
where $k, l$ represent rows of the matrix and $j$ a column of the matrix. Also, $p$ is an integer which determines the vector space where the feature vectors lie [79].

For $p=2$ the Euclidean metric distance is obtained while, for $p=1$ the City Block metric distance.

The disadvantage of the aforementioned metric technique is that the degree of variance amongst the columns of the matrix is not considered. Consequently, coordinates that have high absolute values dominate the result.

The standardized Euclidean metric is defined as:

$$d = \sqrt{(m_k - m_l)D^{-1}(m_k - m_l)^T} \quad (4.25)$$

where $(.)^T$ represents the transpose operator and $D^{-1}$ represents the inverse diagonal matrix. The diagonal elements of $D$ are the variances of each column of the $m$-by-$n$ matrix, whereas the remaining elements of the matrix equal to zero. Although the matrix $D$ has the variance of each column on its main diagonal, the degree of variance amongst the columns of the matrix is still not considered.

Finally, the Mahalanobis metric is defined as:

$$d = \sqrt{(m_k - m_l)V^{-1}(m_k - m_l)^T} \quad (4.26)$$

where $V^{-1}$ represents the inverse covariance matrix. Due to the use of the covariance matrix, the degree of variance amongst the columns of the matrix is considered [78].
Hence, certain coordinate(s) of the feature vectors that are characterized by higher magnitude(s) compared to the rest, do not dictate the classification result.

From the aforementioned metric distance techniques, the Mahalanobis was selected in order to calculate the distance between the test and the target vectors (chapter seven), due to its normalization property. Hence, the classifier used for the audio classification application (chapter seven) is the Mahalanobis metric distance.

4.7 Summary

In this chapter, the justification of certain choices made to the audio classification application (chapter seven), were discussed. The issues of sampling and windowing were highlighted in the first two sections while, sections 4.4 and 4.6 were focused on the issue of statistical feature extraction and on the concept of metric classification, respectively. For the audio classification application (chapter seven) each audio recording has to be divided into frames (windowing, section 4.2) and then statistical features (section 4.4) are extracted from each of the time – frequency signal representations so as to form the feature vectors introduced to the Mahalanobis classifier (section 4.6). Moreover, in section 4.5, the use of ‘difference’ to the phase related spectrograms, is described. This issue will also be discussed in chapter seven (audio classification application). Finally, section 4.3 was dedicated to the issue of finding the ‘roots’ of high-order polynomials that is related to the implementation of the phase spectrum of a signal based on the z-transform, which will be discussed in chapter five.
CHAPTER 5
PHASE ANALYSIS AND PRELIMINARY RESULTS

Introduction

The Fourier phase function of non-synthetic signals is highly discontinuous ([41] and subsection 5.1.3). In this chapter, the two categories of phase discontinuities that appear in the Fourier phase spectrum will be examined. The first category of discontinuities is related to the use of the inverse tangent function while, the second category is related to properties of the signal. Methods to overcome both categories of phase discontinuities are studied. Moreover, in this chapter, the Hartley phase spectrum is explored and its attractive properties, compared to its Fourier counterpart, are observed. Finally, the behavior of the Fourier phase function will be compared to the behavior of the Hartley phase function, via their corresponding cepstrums, in a noise-free (chapter five) and in a noisy (chapter six) environment via synthetic signals.

The discontinuities that the Fourier phase function exhibits cause ambiguities to any feature extraction method applied to the function (section 4.4, chapter four). Thus, by
overcoming the discontinuities of the phase function, the ambiguities caused by their existence, are reduced. Hence, in our research it is shown that in the case where the discontinuities of the phase spectrum are compensated / removed then, one can extract useful information for various applications.

In the conventional phase spectrum, the majority of the discontinuities are introduced due to the use of the inverse tangent function. Hence, the first category of discontinuities is related to the artifacts caused because of the use of the inverse tangent function and not due to structural (i.e., intrinsic) properties of the signal. Consequently, this category of discontinuity will be referred to as ‘extrinsic’, named after the source that causes it. The second category of discontinuity appears when both the real and the imaginary components of the Fourier spectrum at a frequency point of the signal become zero, simultaneously. Equivalently, in the z-domain, the same category of discontinuity appears when a ‘zero’ ('root') of the signal is located on the circumference of the unit circle [39]. Consequently, this category of discontinuity is caused by the structure of the signal and will be called ‘intrinsic’, named after the source that causes it.

Hence, this chapter is organized in six sections. The first section (section 5.1) provides the theoretical analysis of the phase estimation based on the Fourier transform. The two categories of discontinuities (‘extrinsic’ and ‘intrinsic’) that appear in the Fourier phase spectrum are introduced, as well as the methods to overcome them. The second section of this chapter (section 5.2) provides examples that justify the theoretical concepts analyzed in the first part. Subsequently, the third section (section 5.3) provides the theoretical analysis of the phase estimation based on the Hartley transform (Hartley phase spectrum / ‘whitened’ Hartley spectrum) as well as its advantages compared to the phase estimation based on the Fourier transform. The fourth section (section 5.4) demonstrates the attractive properties of the implementation of the phase spectrum based on the Hartley rather than the Fourier approach via examples of synthetic signals. Note that, in both the experimental sections (sections 5.2 and 5.4) the examples provided are in a noise-free environment. The superiority of the Hartley phase spectrum (‘whitened’ Hartley spectrum), compared to its Fourier counterpart, in a noisy environment is demonstrated
via examples in the following chapter (chapter six). Moreover, in section 5.5, a step-by-step description of all six phase related algorithms employed for the time localization of the signal, based on the phase function (chapter six) and the audio classification (chapter seven) applications, is given. Finally, in section 5.6 a summary of this chapter, is provided.

5.1 Phase spectrum implementation based on the Fourier transform

In this section, the method for the evaluation of the phase spectrum of a signal, will be presented. Initially, the signal is transformed to the frequency domain via the DTFT (Discrete-Time Fourier Transform). The Fourier phase spectrum is constructed based on the following formula:

$$\phi(\omega) = \arctan \left( \frac{\Im(S(\omega))}{\Re(S(\omega))} \right)$$

where $$-\pi < \phi(\omega) < \pi$$, \(S(\omega)\) is the complex Fourier spectrum of the signal and \(\Im(.)\), \(\Re(.)\) are its imaginary and real components, respectively.

During computation, the range of the values of the inverse tangent function is restricted to \(2\pi\). Hence, an 'extrinsic' discontinuity is said to appear in the continuous phase spectrum whenever the difference between two consecutive phase angles is \(2\pi\) radians. However, since the signals employed are discrete, an 'extrinsic' discontinuity is said to appear in the discrete phase spectrum, whenever the difference between two consecutive phase samples is higher than a certain threshold [41]. This occurs because the sampling rate (subsection 4.1.2, chapter four) is not always adequate to detect the exact value of the phase transition between rapidly changing phase samples.

The phase spectrum of a signal, without the removal of the 'extrinsic' discontinuities, is characterized by rapid phase transitions. Hence, the phase spectrum of a signal without the removal of the 'extrinsic' discontinuities seems like a noisy signal. Consequently, it is not possible to derive any features from this highly discontinuous function. Thus, the
removal of the ‘extrinsic’ discontinuities is a necessary process. The method used in order to compensate the ‘extrinsic’ discontinuities is presented in subsection 5.1.1.

The second category of discontinuities appears when both the real and the imaginary components of the signal’s complex spectrum become zero simultaneously. In that case the phase value for this particular frequency point (equation (5.1)) tends to $\arctan \left( \frac{0}{0} \right)$ [39]. Note that the ‘intrinsic’ discontinuities, unlike the ‘extrinsic’ discontinuities, form part of the signal and consequently their compensation/removal results in information distortion/loss. Moreover, as will be shown in chapter six, the removal of the ‘intrinsic’ discontinuities is not always necessary. Thus, when dealing with time localization related problems, the removal of the ‘intrinsic’ discontinuities can be omitted (chapter six). However, when the application involves dimensionality reduction of non-artificial signals, such as classification of audio utterances based on statistical feature extraction, their compensation/removal provides better results (chapter seven), in terms of classification rate. As it will be presented in subsections 5.1.1, 5.1.2 and 5.1.3, the ‘extrinsic’ discontinuities can be compensated whereas, the ‘intrinsic’ discontinuities can be either compensated or removed.

5.1.1 Compensation of the ‘extrinsic’ discontinuities

The ‘extrinsic’ discontinuities are compensated by adding $\pm 2\pi$ radians, whenever there is a phase transition greater than $\pi$ radians between two consecutive phase angles [41], in the signal’s phase spectrum. The disadvantage of this method is that the identification of the discontinuities is made difficult due to the discrete nature of the phase function which can change rapidly with frequency. Hence, the ‘unwrapping’ algorithm cannot discriminate between ‘wrapping’ ambiguities (i.e., artifacts caused due to the use of the inverse tangent function) and rapidly changing phase angles (i.e., structural features of the signal). An overview of the phase ‘unwrapping’ algorithms, based on [41], is presented in section 2.3 of chapter two. Finally, note that, unlike the case of the ‘intrinsic’
discontinuities, there is not a known method in order to remove, rather than just to compensate, the ‘extrinsic’ discontinuities.

The second category of discontinuities (‘intrinsic’) can be either compensated or removed. In subsection 5.1.2, the compensation of the ‘intrinsic’ discontinuities will be discussed whereas, in subsection 5.1.3 the removal of the ‘intrinsic’ discontinuities, will be presented.

5.1.2 Compensation of the ‘intrinsic’ discontinuities (via the DTFT)

Whenever there is a simultaneous change of signs of the real and the imaginary components of two subsequent frequency points, then there is a simultaneous zero crossing in the region between them. These frequency points will be referred to as ‘critical’ points, whilst the region between these two points (region where the zero crossing occurs) will be referred to as ‘critical’ region. Note that, the signs of the real and the imaginary Fourier components define the quadrant in which the corresponding Fourier phase angle falls. When both components are positive the angle lies in the first quadrant \((0 \rightarrow \frac{\pi}{2} \text{ radians})\) while, when they are both negative the phase angle lies in the third quadrant \((\pi \rightarrow \frac{3\pi}{2} \text{ radians})\). When the real component is positive and the imaginary component is negative the phase angle lies in the fourth quadrant \((\frac{3\pi}{2} \rightarrow 2\pi \text{ radians})\) and when the real component is negative and the imaginary component is positive the phase angle falls in the second quadrant \((\frac{\pi}{2} \rightarrow \pi \text{ radians})\). Thus, a simultaneous change of signs of the real and the imaginary components of two subsequent frequency points (i.e. zero crossing), is equivalent to a phase transition from either the first to the third or the second to the fourth quadrant or vice versa, in the phase spectrum. Therefore, these phase transitions are approximately equal to \(\pi\). Note that, after the compensation of the
'extrinsic' discontinuities, the 'intrinsic' discontinuities cause the highest transitions in the phase spectrum.

Thus, the compensation rule for the 'intrinsic' discontinuities states: starting from the beginning of the phase spectrum (i.e., 0 frequency point), if the phase angle of a 'critical' point is greater compared to the phase angle of the 'critical' point before it, then \(-\pi\) is added to the rest of the spectrum and vice versa. Consequently, the addition/subtraction of \(\pi\) radians reduces the phase transitions by changing the quadrant of the frequency points. From now on, this method of compensation of the 'intrinsic' discontinuities will be called phase evaluation via the DTFT with compensation of the 'intrinsic' discontinuities.

Hence, the phase evaluation process with compensation of the 'intrinsic' discontinuities can be summarized in three stages. The first stage is the evaluation of the phase spectrum based on equation (5.1), the second stage is the application of the 'unwrapping' algorithm (compensation of the 'extrinsic' discontinuities, subsection 5.1.1) to the phase spectrum, while the last stage is the application of the aforementioned compensation algorithm.

The necessity for the compensation of the 'intrinsic' discontinuities is demonstrated via an example with a synthetic signal. Consider a 256-samples long signal comprised of multiple excitations. The 26 equally spaced unit samples, are repeating themselves every 10 samples whereas, the remaining samples are fixed to zero. Figure 5.1a shows the signal in the time domain, figure 5.1b shows the magnitude spectrum of the signal and figures 5.1c and 5.1d show how the real and imaginary components, of the signal, evolve in the spectrum.
Figure 5.1a: Multiple excitation signal
Figure 5.1b: Magnitude spectrum of the multiple excitation signal
Figure 5.1c: Real part of the multiple pulses
Figure 5.1d: Imaginary part of the multiple pulses

Figure 5.2a presents the 'unwrapped' phase spectrum (subsection 5.1.1) of the signal presented in figure 5.1a. It is important to mention that even in the case where the 'unwrapping' algorithm would not be applied to the phase spectrum of the aforementioned signal, figure 5.2a would be the same. This occurs because the phase transitions in the phase spectrum are lower than $\pi$ and consequently, the 'unwrapping' algorithm cannot correct them. However, applying the detection and compensation rule of the 'intrinsic' discontinuities, figure 5.2b, is obtained.
Chapter 5. Phase analysis and preliminary results

Figure 5.2a: Fourier phase spectrum — removal of the ‘extrinsic’ discontinuities (upper figure)

Figure 5.2b: Fourier phase spectrum — removal of the ‘extrinsic’ and the ‘intrinsic’ discontinuities (lower figure)

One can observe from figure 5.2b, that all the discontinuities are detected and removed from the phase spectrum. The small peaks that appear around the 60th and before the 200th (symmetric) frequency samples (figure 5.2b), are due to the approximation related ambiguities. From figures 5.1c, 5.1d and 5.2a, one can observe that the ‘intrinsic’ discontinuities occur when there is a simultaneous zero crossing of the real and the imaginary components of the signal (i.e. when a spectral peak appears in the same frequency point in both the real (figure 5.1c) and the imaginary components (figure 5.1d) of the signal - ‘intrinsic’ discontinuity.

A step-by-step description of the implementation of the Fourier phase spectrum via the Discrete-Time Fourier Transform (DTFT) is provided in section 5.5 (algorithm 4).
5.1.3 Removal of the 'intrinsic' discontinuities (via the z-transform)

In the previous subsection, the compensation of the 'intrinsic' discontinuities was discussed whereas, in this subsection the removal of the 'intrinsic' discontinuities, is presented. Hence, the method in order to remove the 'intrinsic' discontinuities can be summarized as follows.

Consider a sequence of discrete-time data points: \[ s = [s_0, s_1, ..., s_M] \] with \( s_0 \neq 0 \). The discrete-time data points can be expressed in the z-domain as \( S(z) = \sum_{k=0}^{M} s_k z^{-k} \) (5.3), where \( z \) is the z-plane complex variable. Since \( S(z) \) is a polynomial in \( z \), it can be expressed in factored form as:

\[ S(z) = s_0 z^{-M} \prod_{k=1}^{M} (z - z_k) \] (5.4), and each value of \( z_k \) represents each 'zero' in the z-plane. The unit circle is divided into \( N \) discrete frequency points. Thus, the phase value of a single 'zero' \( z_k \), with respect to a single frequency point \( \omega \), of the unit circle, is equal to:

\[ \phi_{\omega k} = \arctan \left( \frac{\Im(z_k) - \sin(\omega)}{\Re(z_k) - \cos(\omega)} \right) \text{ where } 0 \leq \phi_{\omega k} < 2\pi \] (5.5)

and thus, the phase for all \( M \) 'zeros' in the z-plane, with respect to the same frequency point, \( \omega \), is equal to:

\[ \phi_{\omega T} = \sum_{k=1}^{M} \phi_{\omega k} \] (5.6)

where, \( \Im(z_k) \) and \( \Re(z_k) \) are the imaginary and the real components of each 'zero' \( z_k \).

Thus, in order to evaluate the phase spectrum of the signal, the aforementioned process is repeated for each frequency point \( \omega \) (equation (5.6)) that the unit circle is divided into and consequently,

\[ \phi_{\text{TOTAL}} = [\phi_{1\omega}, \phi_{2\omega}, ..., \phi_{\omega T}, ..., \phi_{M\omega}] \] (5.7)
A schematic representation of the aforesaid algorithm for a single ‘zero’ in the z-plane follows in figure 5.3.

Figure 5.3: Calculation of phase response of a ‘zero’ in the z-plane (taken from [66], pages 321 and 322)

The phase contribution from each ‘zero’, \( z \), is evaluated with respect to each frequency point, \( C \) (forms angle \( w \) with the positive real axis), into which the unit circle is divided [86]. In figure 5.3, for simplicity, the phase spectrum of the ‘zero’ \( z \) is evaluated with respect to a single frequency point only \( C \).

Thus, for the Fourier case (from equation (5.5)), the phase value of a ‘zero’ \( z \) with respect to a frequency point located in \( C \) is given by:

\[
\varphi(w) = \arctan \left( \frac{\Im(z) - \sin(w)}{\Re(z) - \cos(w)} \right) = \arctan \left( \frac{BC}{AB} \right) \quad (5.8)
\]

where \( 0 < \varphi(w) < 2\pi \) and \( \Im(.) \), \( \Re(.) \) are the imaginary and the real components of the ‘zero’ \( z \), respectively. For the case where a signal consists of more ‘zeros’, the phase contributions from each ‘zero’, with respect to a single frequency point, are summed, i.e.
\[ \varphi_{\text{TOTAL}} = \sum_{\ell=1}^{M} \varphi(\nu) \quad (5.9) \]

for the \( M \) 'zeros' located in the z-plane [86].

The same process is repeated for each frequency point and in that way, the Fourier phase spectrum via the z-transform is constructed.

Note that, 'extrinsic' discontinuities do not exist in the phase spectrum derived via the z-transform (unlike the evaluation of the phase spectrum via the DTFT). In this case, although each phase component still uses the inverse tangent function, its value is constrained to \( \pm \pi \) radians and consequently, no 'extrinsic' discontinuities arise, i.e., phase is not 'wrapped' around zero.

However, the 'intrinsic' discontinuities are still present in the phase spectrum. The 'intrinsic' discontinuities are caused by the 'zeros' that are located on the circumference of the unit circle. Assume that the 'zero' (\( z \)) of figure 5.3 is located on the circumference of the unit circle and coincides with the frequency point \( C \) then, equation (5.8) tends to

\[ \varphi(\nu) = \arctan \left( \frac{0}{0} \right) \]

and hence, the 'intrinsic' discontinuity appears \( (AB = BC = 0, \) (triangle \( ABC \)). Consequently, removing the 'zeros' that are located on the circumference of the unit circle and constructing the phase spectrum from the remaining 'zeros' results in a phase spectrum without 'intrinsic' discontinuities. Thus, a 'ring' is drawn around the circumference of the unit circle and all the 'zeros' located within that 'ring' are removed. Then, the phase spectrum is constructed from the remaining 'zeros'. The width of the 'ring' has to be kept within certain limits, in order to keep information loss to a minimum.

Moreover note that the 'zeros' located very close to the circumference of the unit circle cause rapid phase angle transitions in the phase spectrum [39]. These rapid phase angle transitions are not discontinuities, however, for certain applications they affect the performance of the system. As will be shown in chapter seven, both the 'zeros' located on the circumference of the unit circle as well as the 'zeros' located very close to it, affect
the classification rate. Hence, the removal of the 'zeros' located on the circumference of
the unit circle as well as the removal of a certain number of 'zeros' located within a
certain distance from its circumference, evaluating the phase spectrum with the remaining
'zeros', improves the performance of the system, in terms of classification. From now on,
this method of compensation of the 'intrinsic' discontinuities will be called phase
evaluation via the z-transform with removal of the 'intrinsic' discontinuities.

A step-by-step description of the implementation of the Fourier phase spectrum via the z-
transform is provided in section 5.5 (algorithm 1).

The majority of the 'zeros' of the non-synthetic audio signals (e.g., speech, mechanical
sounds, etc.), are located on, or very close to the circumference of the unit circle. In the
following part an example is provided, that presents the location of the 'zeros' in the z-
plane of a non-synthetic signal (gunshot) taken from our database. Details related to the
database used are provided in chapter seven. Hence, figure 5.4, shows the location of the
'zeros' of a 256 sample frame of an utterance from class (vii) (chapter seven), in the z-
plane. The circular markers denote the 'zeros' location.

Figure 5.4: ‘Zeros’ located in the z-plane
(of a single 256 samples long frame from
class (vii))

Figure 5.5: ‘Zeros’ located in the z-plane
(of a single – same as in figure 5.4 - 256
samples long frame from class (vii), after
the application of the exclusion ‘ring’ with
width 0.004)
Figure 5.5 shows the 'zeros' that remain after the removal of the 'zeros' located within a 0.004 distance from the circumference of the unit circle. The effect of the application of the exclusion 'ring' is the removal of the 'intrinsic' discontinuities ('zeros' locate on the unit circle) and the reduction of the number of the rapidly changing phase angles ('zeros' locate close to the unit circle) from the phase spectrum. However, note that there is a trade-off between the amount of information loss and the reduction in the number of the discontinuities in the phase spectrum. In chapter seven, using the methods described here, phase spectrograms are employed as part of the feature set and the effect of the width of the exclusion 'ring' on the classification rate is also investigated.

5.1.4 Summary of the methods for the compensation/removal of the 'extrinsic'/intrinsic' discontinuities

In subsections 5.1.1 to 5.1.3, the method for the compensation of the 'extrinsic' discontinuities, the method for the compensation of the 'intrinsic' discontinuities and the method for the removal of the 'intrinsic' discontinuities, were presented. Hence, as mentioned, the method for the removal of the 'extrinsic' discontinuities, from the phase spectrum, is based on an heuristic approach of detecting and compensating the phase transitions that are higher than $\pi$.

The advantage of the implementation of the phase spectrum via the DTFT with the compensation of the 'intrinsic' discontinuities (subsection 5.1.2), compared to the method of implementation of the phase spectrum via the z-transform with the removal of the 'intrinsic' discontinuities (subsection 5.1.3), is that all the data points are used for the construction of the phase spectrum.

However, the phase evaluation via the z-transform (subsection 5.1.3) is characterized by two advantages compared to the phase evaluation via the DTFT (subsections 5.1.1 and 5.1.2). The first advantage is the lack of 'extrinsic' discontinuities, while the second advantage is that there is not a similar straightforward method for the removal (not compensation) of the 'intrinsic' discontinuities based on the DTFT approach. The major
disadvantage of the phase evaluation via the z-transform is the uncertainties introduced
due to the mislocation of the 'roots' on the unit circle. Hence, in order to keep these
inaccuracies to low levels, the polynomial size is limited to 256 samples only (i.e., the
signal is divided in frames of 256 samples each – chapter seven).
In section 5.3, the implementation of the phase spectrum via the Hartley transform will be
discussed, as well as its advantages over its Fourier counterpart.

5.2 Experimental results of synthetic signals based on the DTFT
approach

In this section, two examples of the time localization of the signal, based on the phase
function implemented via the DTFT (subsection 5.1.2), will be presented. These two
examples are provided so as to show that the phase function encapsulates the time
localization related information of a signal and also to demonstrate the effect of the
removal of the 'intrinsic' discontinuities from the phase spectrum. The first example uses
an exponential decaying sine wave as the test signal whereas, the second example uses
the rectangular pulse as the test signal.

The 'zeros' location of an exponentially decaying sine wave is shown in figure 5.6. Note
that there are no 'zeros' ('roots' of the polynomial) located on the circumference of the
unit circle and consequently, the 'intrinsic' category of discontinuities is not present in its
phase spectrum. Assume that the exponentially decaying sine wave is located from
sample 251 to sample 280 of the 2048 samples long time-domain signal (i.e., the time
domain samples from 1 to 250 and from 281 to 2048 are set to zero). Figure 5.6 presents
the location of the 'zeros' of the aforementioned exponentially decaying sine wave in the
z-plane. Hence, from figure 5.6, note that the total number of 'roots' appearing in the z-
plane are: 1769 (origin) + 27 (close to the circumference of the unit circle) = 1796. This
is explained as: a signal that is 2048 samples long should form a polynomial of the same
power. However, the starting point of the signal is the 251st sample and consequently, the
first non-zero coefficient of the polynomial appears in the same sample. Hence, in theory
the number of the 'roots' appearing in the z-plane should be 1797. The one missing
'root', i.e. 1797 (theoretical number of 'roots') - 1796 (number of 'roots' appearing in the
z-plane), occurs because its value is very small and consequently, it does not appear on
the z-plane, although it exists and participates in the process (figure 5.6).

![Zeros location in the z-plane of an exponentially decaying sine wave](image1)
![Zeros location in the z-plane of a rectangular pulse](image2)

Figure 5.6: ‘Zeros’ located in the z-plane of an exponentially decaying sine wave (located from 251 to 280)
Figure 5.7: ‘Zeros’ located in the z-plane of a rectangular pulse (located from 251 to 280, figure 5.11a)

The second signal chosen is a rectangular pulse (figure 5.7). Assume that the rectangular
pulse is located from the 251st sample to the 280th sample of the 2048 samples long time
domain signal, with its amplitude equal to unity. This signal is chosen because its phase
spectrum, unlike the phase spectrum of the exponentially decaying sine wave, conveys
‘intrinsic’ discontinuities. The existence of the ‘intrinsic’ discontinuities is indicated by
the location of ‘roots’ on the circumference of the unit circle (figure 5.7). Moreover, for
the case of the rectangular pulse, the total number of ‘roots’ appearing in the z-plane
should be equal to the theoretical number of ‘roots’ of the exponential decaying sine
wave (figure 5.6), because both of the signals have the same starting point in the time
domain (251st sample). So, the total number of ‘roots’ is 1768 (origin) + 29 (on the
circumference of the unit circle) = 1797, as expected (figure 5.7).

For time localization related problems, the implementation of the phase spectrum via the
DTFT with compensation of the ‘intrinsic’ discontinuities (subsection 5.1.2) is preferred
to the implementation of the phase spectrum via the z-transform with removal of the
‘intrinsic’ discontinuities (subsection 5.1.3) because synthetic signals, such as the
rectangular pulse, have some of its ‘zeros’ located on the circumference of the unit circle and thus, the ‘zeros’ removal approach discards important information related to the shape of the signal. Moreover, the z-transform approach introduces ambiguities due to the mislocation of the ‘roots’ on the unit circle (subsection 5.1.4).

In order to express the phase content of a signal in a useful manner, the homomorphic deconvolution process (section 3.2, chapter three), is used. As a reminder, for the Fourier case, the first, the second and the third stages of figure 5.8, are the DTFT, the evaluation of the Fourier phase spectrum and the IDTFT, respectively.

The following figure summarizes the homomorphic deconvolution process.

![Figure 5.8: Summary of the homomorphic deconvolution process](image)

The non-linear process (i.e. phase spectrum implementation via the DTFT) (2\textsuperscript{nd} stage of figure 5.8), is based on the method described in subsection 5.1.2, i.e., phase evaluation via the DTFT with compensation of the ‘intrinsic’ discontinuities. Thus, for the Fourier case, the non-linear process (2\textsuperscript{nd} stage of figure 5.8) can be divided into the three stages:

![Figure 5.9: Stages of the non-linear part of the homomorphic deconvolution process applied to the Fourier case](image)

Figures 5.10a, 5.11a and 5.12a show the location of three rectangular pulses, of the same shape, in the time domain. Figures 5.10b, 5.11b and 5.12b, illustrate the effect that time shifting has on the Fourier phase function (2\textsuperscript{nd} stage of the homomorphic deconvolution...
process, figure 5.8), for the same rectangular pulses (figures 5.10a, 5.11a and 5.12a). As can be seen, the Fourier phase spectrum is a ramp function [112]. The gradient of this ramp depends on the location of the signal in the time domain. Applying the IDTFT to this ramp, results in the representation of the signal in the cepstral domain (figures 5.10c, 5.11c and 5.12c) (i.e., 3rd stage of the homomorphic deconvolution process, figure 5.8). The maximum value of the cepstral function, corresponds to the location of the pulse in the time domain, e.g., if the maximum value in the cepstral domain equals to 81 it means that the signal is located around the 81st point in the time domain. This point does not always correspond to the first point of the signal in the time domain, but depending on the shape of the pulse and its amount of shift in time, it could correspond to one of the initial points (second, third) or even the middle point in case where the original signal is characterized by a dominant lobe e.g. a gaussian pulse.

As it will be shown in the following part, the amplitude of the cepstral function has a linear relationship with the gradient of the Fourier phase function (after ‘unwrapping’ is applied) and consequently, corresponds to the location of the signal in the time domain.

Consider the case of a dirac-delta pulse in the time domain. Assuming that $\varphi(\omega)$ is the Fourier phase function (2nd stage of figure 5.8) then,

$$\varphi(\omega) = -k\omega \quad (5.10)$$

where $k$ corresponds to the gradient of the Fourier phase function. Consequently, the Fourier phase cepstrum (IDTFT of the Fourier phase function - 3rd stage of figure 5.8), $c_\varphi(\tau)$, is expressed as:

$$c_\varphi(\tau) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi(\omega)e^{j\omega\tau} d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} (-k\omega) e^{j\omega\tau} d\omega = -\frac{k}{2\pi} \int_{-\pi}^{\pi} \omega e^{j\omega\tau} d\omega$$
Hence, the amplitude of the cepstral peak in the Fourier phase cepstrum, \( c_F(\tau) \), depends on the gradient of the Fourier phase function \( k \).

Thus, the amplitude of the Fourier phase cepstrum, \( c_F(\tau) \) (figure 5.10c), forms a linear relationship with \( k \) (figure 5.10b) that depends on the location of the signal in the time domain (figure 5.10a). Note that, due to the shape of the Fourier phase function (ramp) (figure 5.10b), which is odd, the cepstral information is encapsulated in the imaginary part of the IDTFT (3rd stage of figure 5.8) of the Fourier phase function.

In order to reduce the ambiguities due to aliasing artifacts, all the signals were shifted only between the 1st and the 512th out of the 2048 points of the time domain, i.e., a quarter of the total length available.

Table 5.1 presents the amplitude of the 0th cepstral value in the Fourier phase cepstral domain (i.e. 3rd stage of the homomorphic deconvolution process, figure 5.8) for an exponential decaying sine wave (2nd and 3rd columns) and a rectangular pulse (4th and 5th columns), as the pulses shift in the time domain (1st column).
For the case of the exponentially decaying sine wave (2nd and 3rd columns of table 5.1), the amplitude of the 0th cepstral value, is the same, irrespective of the removal (3rd column) or not (2nd column) of the 'intrinsic' discontinuities. As mentioned, this occurs because the exponentially decaying sine wave has no 'zeros' located on the circumference of the unit circle and consequently, no 'intrinsic' discontinuities arise. Moreover, note that for both cases of the test signals (exponentially decaying sine wave and rectangular pulse) the amplitude of the 0th cepstral value corresponds to the location of the pulse in the time domain (1st column).

In terms of shape, the Fourier phase spectrum and the Fourier phase cepstrum after the removal of both the ‘extrinsic’ and the ‘intrinsic’ discontinuities are similar compared to the Fourier phase spectrum and the Fourier phase cepstrum after the removal of the ‘extrinsic’ discontinuities only. The only difference is a slight change in the gradient of the Fourier phase spectrum and consequently, a similarly slight change in the amplitude of the 0th cepstral value, in the cepstral domain (4th and 5th columns of table 5.1).
effect of the removal of the 'intrinsic' discontinuities, is more obvious for the Hartley phase spectrum case as will be presented in section 5.4 and chapter six. Moreover, examples of the behavior of the Fourier phase function in a noisy-environment will be provided in chapter six.

Note that, the homomorphic deconvolution process (figure 5.8) requires invertibility. However, the second stage of the non-linear process (figures 5.8 and 5.9) is not invertible, due to the heuristic manner in which the 'unwrapping' algorithm compensates the 'extrinsic' discontinuities (subsection 5.1.1). For simple cases though, such as the implementation of the phase spectrum/cepstrum of a single pulse in a noise free environment, the 'unwrapping' algorithm does not affect the process. As will be shown in chapter six, the ambiguities introduced by the 'unwrapping' algorithm is the reason that the Fourier phase cepstrum cannot be used, as a time localization technique, for more complicated cases, such as to specify the position of more than a single pulse in a noisy environment.
Chapter 5. Phase analysis and preliminary results

Figure 5.10a: Time domain representation of a rectangular pulse
Figure 5.10b: Frequency (Fourier phase spectrum) domain representation of a rectangular pulse
Figure 5.10c: Cepstral (Fourier phase cepstrum) domain representation of a rectangular pulse
Figure 5.11a: Time domain representation of a rectangular pulse
Figure 5.11b: Frequency (Fourier phase spectrum) domain representation of a rectangular pulse
Figure 5.11c: Cepstral (Fourier phase cepstrum) domain representation of a rectangular pulse
Chapter 5. Phase analysis and preliminary results

Figure 5.12a: Time domain representation of a rectangular pulse

Figure 5.12b: Frequency (Fourier phase spectrum) domain representation of a rectangular pulse

Figure 5.12c: Cepstral (Fourier phase cepstrum) domain representation of a rectangular pulse
5.3 Phase spectrum implementation based on the Hartley transform

In this section, the methods for the compensation/removal of the 'intrinsic' discontinuities (subsections 5.1.2 and 5.1.3), will be adjusted for the Hartley case. Examples of synthetic signals (exponentially decaying sine wave and rectangular pulse), will demonstrate the advantages of the Hartley compared to its Fourier counterpart.

As mentioned in subsection 3.1.1 of chapter three, the 'scaled' or the 'whitened' Hartley spectrum is defined as;

$$Y(\omega) = \frac{H[s(t)]}{M(\omega)} = \cos(\phi(\omega)) + \sin(\phi(\omega))$$

Equation (5.12) is a function of the Fourier phase, $\phi(\omega)$, only. From now on, $Y(\omega)$ (equation (5.12)), will be called the 'whitened' Hartley spectrum, the Hartley phase spectrum or just the Hartley phase.

As mentioned in the previous sections, one major disadvantage of the phase analysis based on the conventional Fourier phase spectrum is the use of the inverse tangent function, which is a highly discontinuous function. The Hartley phase function (equation (5.12)) does not convey the 'extrinsic' discontinuities. However, the 'intrinsic' discontinuities are present in the function. In subsection 5.3.1, the implementation of the Hartley phase spectrum via the DTHT (Discrete-Time Hartley Transform) is analyzed whilst, in section 5.3.2 the implementation of the Hartley phase spectrum via the z-transform is discussed.

5.3.1 Hartley phase spectrum implementation via the DTHT

In this subsection, the method described in subsection 5.1.2 (implementation of the Fourier phase spectrum via the DTFT), will be developed based on the definition of the
Chapter 5. Phase analysis and preliminary results

Hartley phase (equation (5.12)). As mentioned, the Hartley phase function, as for the Fourier phase function, conveys ‘intrinsic’ discontinuities.

Hence, based on the method described in subsection 5.1.2, these discontinuities may be removed in the following way:

assume that ‘b’ represents the phase value at a ‘critical’ point and then for the Fourier case, assume that π has to be added for compensation.

Hence, based on equation (5.12), the Hartley phase for ‘b’ is:

\[ \cos(b + \pi) + \sin(b + \pi) = - (\cos(b) + \sin(b)) \] after compensation

and if \( \pi \) has to be subtracted then,

\[ \cos(b - \pi) + \sin(b - \pi) = - (\cos(b) + \sin(b)), \] which is the same as before.

Thus, in either case, the compensation for an ‘intrinsic’ discontinuity is achieved by multiplication of the Hartley phase function with -1. Therefore, after the ‘critical’ points have been detected, the Hartley phase spectrum is then scanned from the start (0 frequency point). The rest of the Hartley phase spectrum is multiplied with (-1) either if the value before the ‘critical’ point is higher or if it is lower.

A step-by-step description of the implementation of the Hartley phase spectrum via the Discrete-Time Hartley Transform (DTHT) is provided in section 5.5 (algorithm 5).

Concluding, one advantage of the phase evaluation via the DTHT compared to the phase evaluation via the DTFT is that the Hartley phase does not convey the ‘extrinsic’ discontinuities (no use of the inverse tangent function). Examples of time localization based on the Hartley phase spectrum implementation via the DTHT, are presented in section 5.4.
5.3.2 Hartley phase spectrum implementation via the z-transform

The method described in subsection 5.1.3, may also be adapted for use with the Hartley phase. The Hartley phase may be evaluated geometrically, (figure 5.13), in the same manner as for Fourier phase.

![Diagram of phase response in the z-plane](image)

Figure 5.13: Phase response of a 'zero' in the z-plane

The phase contribution from each 'zero', \( z \), is evaluated with respect to each frequency point, \( w \), in which the unit circle is divided. For the case shown in figure 5.13, for simplicity, the phase spectrum is evaluated with respect to a single frequency point only. As a reminder, for the Fourier case, the phase value of \( z \) with respect to \( w \) is given by the angle \( \phi \):

\[
\phi(w) = \arctan \left( \frac{BC}{AB} \right) \quad (5.13).
\]

For the Hartley case, based on equation (5.12),

\[
Y(w) = \cos(\phi(w)) + \sin(\phi(w)) = \frac{AB}{AC} + \frac{BC}{AC} = \frac{(AB + BC)}{AC} \quad (5.14).
\]
Thus, from now on, the Hartley phase spectrum implemented using the method described in subsection 5.3.1, will be called the Hartley phase spectrum via the DTHT (Discrete-time Hartley transform) whereas, the Hartley phase spectrum implemented using the method described in section 5.3.2, will be called the Hartley phase spectrum via the z-transform.

A step-by-step description of the implementation of the Hartley phase spectrum via the z-transform is provided in section 5.5 (algorithm 2).
5.4 Experimental results of synthetic signals based on the DTHT approach

In this section, as with the Fourier case (section 5.2), two examples of time localization via the Hartley cepstrum, will be presented. Again, the exponential decaying sine wave and the rectangular pulse will be used as the test signals. The phase analysis will be based on the phase function derived via the DTHT (subsection 5.3.1).

As with the Fourier case (section 5.2), in order to express the phase content of a signal in a useful manner, the homomorphic deconvolution process (figure 5.8), is used (section 3.2, chapter three). Hence, for the Hartley case, the first, the second and the third stages of figure 5.8, are the DTHT, the evaluation of the Hartley phase spectrum and the IDTHT, respectively.

For the Hartley case, the non-linear process (2nd stage of figure 5.8) can be divided in the following two stages:

Figure 5.14: Stages of the non-linear part of the homomorphic deconvolution process applied to the Hartley case

Specifically, the input to the first stage of figure 5.14 is the real and the imaginary components of the Fourier transform of the signal (section 5.3). Then, the 'whitened' Hartley (Hartley phase) spectrum is evaluated based on equation (5.12). Subsequently, the 'intrinsic' discontinuities of the Hartley phase spectrum are compensated based on the theory stated in subsection 5.3.1.
Note that, unlike the Fourier case (figure 5.9), both stages of the non-linear process of the evaluation of the Hartley phase spectrum with compensation of the 'intrinsic' discontinuities are invertible, because the 'unwrapping' algorithm is not used.

Moreover, unlike the case of the Fourier cepstrum (section 5.2), the phase content of the signal in the Hartley cepstrum is encapsulated in the location of the maximum amplitude of the cepstral peak(s) rather than the amplitude of the 0th cepstral value.

Consider the case of a dirac-delta pulse located at $T$ in the time domain. From equation (5.12), the Hartley phase spectrum of this signal is a cosinusoidal function that encapsulates the Fourier phase information $\varphi(\omega)$. Hence, assuming that $y(\omega)$ is the Hartley phase function (2nd stage of figure 5.8) then,

$$y(\omega) = \cos\left(\frac{2\pi}{T} \omega\right) - \sin\left(\frac{2\pi}{T} \omega\right) \quad (5.15).$$

Consequently, the Hartley phase cepstrum (i.e. IDTHT of the Hartley phase spectrum) (3rd stage of figure 5.8), $c_H(\tau)$, is expressed as:

$$c_H(\tau) = \text{IDTHT}(y(\omega)) = \Re\{\text{DTFT}(y(\omega))\} + \Im\{\text{DTFT}(y(\omega))\} \quad (5.16)$$

However,

$$\text{DTFT}(y(\omega)) = \delta(\tau - T) + \delta(\tau + T) + j(\delta(\tau - T) - \delta(\tau + T)) \quad (5.17)$$

Hence, $c_H(\tau)$ (i.e. equation (5.16)), equals,

$$c_H(\tau) = 2\delta(\tau - T) \quad (5.18)$$
Consequently, the phase content of the signal is encapsulated in the location of the delta function (cepstral peak(s)) in the Hartley phase cepstrum $c_h(\tau)$ and not in its amplitude (Fourier case $c_F(\tau)$ - section 5.2).

Table 5.2 presents the location of the maximum amplitude of the Hartley phase cepstral function (i.e. 3rd stage of the homomorphic deconvolution process, figure 5.8) for an exponential decaying sine wave (2nd and 3rd columns) and a rectangular pulse (4th and 5th columns), as the pulses shift in the time domain (1st column).

<table>
<thead>
<tr>
<th>discrete-time domain (starting &amp; finishing point)</th>
<th>exponential pulse (no added noise)</th>
<th>rectangular pulse (no added noise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cepstral domain – 'intrinsic' discontinuities not removed (location of the maximum amplitude of the peak in the Hartley phase cepstral domain)</td>
<td>cepstral domain – 'intrinsic' discontinuities removed (location of the maximum amplitude of the peak in the Hartley phase cepstral domain)</td>
<td>cepstral domain – 'intrinsic' discontinuities not removed (location of the maximum amplitude of the peak(s) in the Hartley phase cepstral domain)</td>
</tr>
<tr>
<td>51-80</td>
<td>51</td>
<td>51</td>
</tr>
<tr>
<td>151-180</td>
<td>151</td>
<td>151</td>
</tr>
<tr>
<td>251-280</td>
<td>251</td>
<td>251</td>
</tr>
<tr>
<td>351-380</td>
<td>351</td>
<td>351</td>
</tr>
<tr>
<td>451-480</td>
<td>451</td>
<td>451</td>
</tr>
</tbody>
</table>

Table 5.2: Location of the maximum amplitude in the Hartley phase cepstral domain for the exponentially decaying sine wave pulse and for the rectangular pulse.

For the case of the exponentially decaying sine wave, there is only a single dominant peak in the Hartley cepstral domain (2nd and 3rd columns of table 5.2). Moreover, for the exponentially decaying sine wave, the location of the maximum amplitude is the same...
irrespective of the removal (3rd column) or not (2nd column) of the 'intrinsic' discontinuities. As mentioned in section 5.2, this occurs because there are no 'zeros' of the signal located on the circumference of the unit circle and hence, no 'intrinsic' discontinuities exist in the phase spectrum. For the case of the rectangular pulse, there are two dominant peaks in the Hartley cepstral domain, in the case where the 'intrinsic' discontinuities are not removed (4th column). The location of the first peak denotes the starting point of the pulse, while the location of the second its finishing point. However, when the 'intrinsic' discontinuities are removed, there is only a single dominant peak in the Hartley phase cepstrum (5th column), that its location corresponds to a point between the starting and the finishing point of the pulse in the time domain (1st column). Similarly to the Fourier phase cepstrum (section 5.2), for both cases of pulses (exponentially decaying sine wave and rectangular pulse), the Hartley phase cepstrum provides accurate estimation of the location of the signal in the time domain.

Figures 5.15b, 5.16b and 5.17b, present examples of the Hartley phase spectrum whereas, figures 5.15c, 5.16c and 5.17c present examples of the Hartley phase cepstrum, for a rectangular pulse shifted in time (figures 5.15a, 5.16a and 5.17a). Moreover, the figures 5.15b, 5.16b and 5.17b, illustrate the effect that the time shift of a pulse (figures 5.15a, 5.16a and 5.17a) has to the Hartley phase function. The shape of the Hartley phase function is a cosinusoidal signal and the rate of the zero crossings, with respect to the $\omega$-axis, corresponds to the signal location in the time domain. Thus, the higher the rate of the zero crossings of the Hartley phase function, the further the signal is shifted in the time domain. In the cepstral domain (figures 5.15c, 5.16c and 5.17c), two dominant peaks appear corresponding to the location of the signal in the time domain, together with harmonics and other by-products. Hence, the IDTHT of the Hartley phase function (3rd stage of figure 5.8), provides the accurate location of the pulse in the time domain, e.g. if the two highest peaks, in the cepstral domain, appear at points 101 and 123, it is indicated that the pulse has a starting and finishing point at 101 and 123, in the time domain.

From figures 5.15c, 5.16c and 5.17c, it is clear that with the exception of the two dominant peaks, there are also other peaks of lower amplitude across the Hartley phase
cepstrum. The majority of these additional peaks, are the result of the 'intrinsic' discontinuities that, as mentioned before (subsection 5.3.1), exist in the Hartley phase spectrum. These additional peaks can be removed from the Hartley phase cepstrum by compensating the 'intrinsic' discontinuities in the corresponding Hartley phase spectrum (subsection 5.3.1). Figure 5.18a (same as figure 5.15b) presents the Hartley phase function without the removal of its discontinuities, of the pulse presented in figure 5.15a whereas, figure 5.18b presents the Hartley phase function after the removal of the majority of the 'intrinsic' discontinuities (compare figures 5.18a and 5.18b). The compensation of the 'intrinsic' discontinuities in the Hartley phase spectrum (figure 5.18b) reduces the number of these additional peaks appearing in the Hartley phase cepstrum significantly (compare figures 5.15c and 5.18c). However, the phase cepstrum (figure 5.18c), derived via the compensated Hartley phase spectrum (figure 5.18b), provides the center of the pulse rather than both its starting and finishing points (compare figures 5.15c and 5.18c). As will be shown in chapters six and seven, depending on the application, in some cases it is more appropriate to compensate the 'intrinsic' discontinuities, whereas in some other cases it is preferable not to. An additional method recommended in order to reduce the 'intrinsic' discontinuities even further, as well as a discussion of the reason that the method described in subsection 5.3.1 does not remove all the 'intrinsic' discontinuities, is presented in section 5.5 (algorithm five) and in chapter six.

Finally, it is important to mention that there are researchers who claim that it would be more appropriate to extract the phase content of a signal from the 'whitened' Fourier spectrum [115], rather than from the conventional Fourier phase spectrum implemented based on equation (5.1).

The 'whitened' Fourier spectrum is implemented, by dividing the complex Fourier spectrum by the magnitude, for each frequency point. Thus, the 'whitened' Fourier spectrum is defined as:
\[ WF(\omega) = \frac{S(\omega)}{\sqrt{\mathcal{R}^2(S(\omega)) + \mathcal{I}^2(S(\omega))}} \] (5.19)

where, \( S(\omega) \) is the complex Fourier spectrum of the signal and \( \mathcal{R}(\cdot) \), \( \mathcal{I}(\cdot) \) are its imaginary and real components, respectively. A step-by-step description of the implementation of the ‘whitened’ Fourier spectrum is provided in section 5.5 (algorithm 6).

Hence, for the ‘whitened’ Fourier spectrum case, in the second stage of the homomorphic deconvolution process (2\textsuperscript{nd} stage of figure 5.8), the Fourier spectrum of the signal is divided by the magnitude, instead of evaluating the phase function based on the process described in figure 5.9 (conventional Fourier spectrum) or figure 5.14 (Hartley phase spectrum). Subsequently, the ‘whitened’ Fourier cepstrum (3\textsuperscript{rd} stage of figure 5.8), is obtained by evaluating the IDTFT of the ‘whitened’ Fourier spectrum.

An important observation related to the algorithmic implementation of the ‘whitened’ Fourier cepstrum is that its information content is encapsulated in the real part of the IDTFT of the ‘whitened’ Fourier spectrum (equation (5.19)). This occurs because, equation (5.19), is of the form: \( \frac{\text{even} + j(\text{odd})}{\text{even}} \) i.e. \( \text{even} + j(\text{odd}) \) and consequently, the cepstral information is encapsulated in the real part of the IDTFT of the ‘whitened’ Fourier spectrum (equation (5.19)).

In terms of shape, the ‘whitened’ Fourier cepstrum is similar to the shape of the Hartley phase cepstrum. The advantage of the ‘whitened’ Fourier spectrum as compared to the conventional Fourier phase spectrum, is that it encapsulates the signal’s phase content without using the inverse tangent function (i.e., no ‘wrapping’ ambiguities). However, because the ‘whitened’ Fourier spectrum is not using the inverse tangent function, there is not a known method in order to overcome its ‘intrinsic’ discontinuities, such as adding/subtracting \( \pi \) (conventional Fourier spectrum), or multiplying with \(-1\) (Hartley spectrum).
phase spectrum). More details related to the 'whitened' Fourier cepstrum will be provided in chapter six.
Figure 5.15a: Time domain representation of a rectangular pulse
Figure 5.15b: Frequency (Hartley phase spectrum) domain representation of a rectangular pulse
Figure 5.15c: Cepstral (Hartley phase cepstrum) domain representation of a rectangular pulse
Chapter 5. Phase analysis and preliminary results

Time samples
Frequency domain - "whitened" Hartley spectrum - "intrinsic" disc, not removed

Phase samples
Cepstral domain - Hartley phase cepstrum • "intrinsic" disc, not removed

Figure 5.16a: Time domain representation of a rectangular pulse
Figure 5.16b: Frequency (Hartley phase spectrum) domain representation of a rectangular pulse
Figure 5.16c: Cepstral (Hartley phase cepstrum) domain representation of a rectangular pulse
Figure 5.17a: Time domain representation of a rectangular pulse

Figure 5.17b: Frequency (Hartley phase spectrum) domain representation of a rectangular pulse

Figure 5.17c: Cepstral (Hartley phase cepstrum) domain representation of a rectangular pulse
Figure 5.18a: Frequency (Hartley phase spectrum) domain representation of a rectangular pulse – ‘intrinsic’ discontinuities not removed

Figure 5.18b: Frequency (Hartley phase spectrum) domain representation of a rectangular pulse – ‘intrinsic’ discontinuities removed

Figure 5.18c: Cepstral (Hartley phase cepstrum) domain representation of a rectangular pulse – ‘intrinsic’ discontinuities removed
5.5 Step-by-step description of the phase related algorithms

In this section, a step-by-step description of all six phase related algorithms mentioned in this chapter, are provided. These algorithms are tested and compared for the time localization of a signal (chapters five and six) and the classification (chapter seven) applications.

Algorithm 1 - Fourier phase spectrum via the z-transform

Step 1: Consider a sequence of discrete-time data points: \( s = [s_0, s_1, ..., s_M] \) with \( s_0 \neq 0 \).

Step 2: The discrete-time data points can be expressed, in the z-domain as:
\[
S(z) = \sum_{k=0}^{M} s_k z^{-k},
\]
where \( z \) is the z-plane complex variable.

Step 3: Since \( S(z) \) is a polynomial in \( z \), it can be expressed in factored form as:
\[
S(z) = s_0 z^{-M} \prod_{k=1}^{M} (z - z_k),
\]
and each value of \( z_k \) represents each 'zero' in the z-plane.

Step 4: Then, the unit circle is divided into \( N \), frequency points. Thus, the phase of a single 'zero' \( z_k \), with respect to a single frequency point \( w \), equals to:
\[
\phi_{wk} = \arctan \left[ \frac{\Im(z_k) - \sin(w)}{\Re(z_k) - \cos(w)} \right]
\]
where \( 0 \leq \phi_{wk} < 2\pi \) (5.1A)

and thus, the phase for all \( M \) 'zeros' in the z-plane, with respect to the same frequency point, \( w \), equals to:
\[
\phi_{wT} = \sum_{k=1}^{M} \phi_{wk}
\]
(5.2A)

where, \( \Im(z_k) \) and \( \Re(z_k) \) are the imaginary and real component of each 'zero' \( z_k \). Thus, in order to evaluate the phase spectrum of the signal, the aforementioned process is repeated for each frequency point \( w \) that the unit circle is divided and consequently,
\[
\phi_{TOTAL} = [\phi_{1T}, \phi_{2T}, ..., \phi_{wT}, ..., \phi_{NT}]
\]
(5.3A)
Chapter 5. Phase analysis and preliminary results

Removal of the 'intrinsic' discontinuities (optional)

Step 5: An exclusion 'ring' is applied around the circumference of the unit circle in order to remove the 'sharp zeros'. From the 'zeros' left in the z-plane, one can construct a new discrete-time domain signal, \( \hat{s} \), via the inverse of the process. Thus,

\[
\hat{S}(z) = \hat{s}_0 z^{-N} \prod_{i=1}^{N} (z - z_i) \quad (5.4A)
\]

where \( N \leq M \). Then, from equation (5.4A),

\[
\hat{S}(z) = \sum_{i=0}^{N} \hat{s}_i z^{-i}
\]

and subsequently, the new discrete-time signal, evaluated via the inverse z-transform is:

\[
\hat{s} = [\hat{s}_0, \hat{s}_1, \ldots, \hat{s}_N].
\]

Hence, the same process, from step 1 to step 4, has to be repeated with \( \hat{s} \), in order to evaluate the phase spectrum of the new signal.

In practice though, the phase spectrum of the signal, after the application of the exclusion 'ring', can be evaluated directly from the subset or remaining 'zeros' identified by equation (5.4A).

Algorithm 2 - ‘Whitened’ Hartley (Hartley phase) spectrum via the z-transform (1st method)

Step 1: same as Step 1 of Fourier phase spectrum via the z-transform.
Step 2: same as Step 2 of Fourier phase spectrum via the z-transform.
Step 3: same as Step 3 of Fourier phase spectrum via the z-transform.
Step 4: same as Step 4 of Fourier phase spectrum via the z-transform, up to the stage where the result of equation (5.3A) is obtained. Subsequently, based on equation (5.12) (chapter five) that establishes the relationship between the ‘whitened’ Hartley spectrum and the Fourier phase spectrum, the ‘whitened’ Hartley spectrum equals to:
Chapter 5. Phase analysis and preliminary results

\[ Y_{\text{TOTAL}} = \left[ \sin(\phi_{1T}) + \cos(\phi_{1T}), \sin(\phi_{2T}) + \cos(\phi_{2T}), \ldots, \sin(\phi_{NT}) + \cos(\phi_{NT}) \right] \quad (5.5A). \]

Removal of the 'intrinsic' discontinuities (optional)

Step 5: same as Step 5 of Fourier phase spectrum via the z-transform, using again equation (5.5A) rather than equation (5.3A).

Algorithm 3 - 'Whitened' Hartley (Hartley phase) spectrum via the z-transform (2nd method)

Step 1: same as Step 1 of 'whitened' Hartley spectrum via the z-transform (1st method).

Step 2: same as Step 2 of 'whitened' Hartley spectrum via the z-transform (1st method).

Step 3: same as Step 2 of 'whitened' Hartley spectrum via the z-transform (1st method).

Step 4: Similarly to step 4, of 'whitened' Hartley spectrum via the z-transform (1st method), the unit circle is divided into \( N \) frequency points.

As a reminder, the 'whitened' Hartley spectrum (equation (5.12), chapter five), is defined as:

\[ Y(\omega) = \frac{H(\omega)}{M(\omega)} = \cos(\varphi(\omega)) + \sin(\varphi(\omega)) \quad (5.6A), \]

where, \( H(\omega) = \Re(\omega) - \Im(\omega) \) and \( M(\omega) = \sqrt{\Re^2(\omega) + \Im^2(\omega)} \).

In practice, both the 1\textsuperscript{st} and the 2\textsuperscript{nd} methods of implementation of the 'whitened' Hartley spectrum via the z-transform, are based on a 'nested' loop. The 'inner' loop considers the phase contribution of each 'zero', with respect to a single frequency point \( \omega \), whilst the 'outer' loop repeats the same process for each frequency point, in which the unit circle is divided, in order to obtain the phase spectrum of the signal.

As explained in algorithm 2, for the 1\textsuperscript{st} method of implementation of the 'whitened' Hartley spectrum via the z-transform, as with the implementation of the Fourier phase
spectrum via the z-transform, the inner loop *sums* the phase contribution of each ‘zero’ (equation (5.2A)), with respect to a single frequency point \( w \) whereas, the outer loop repeats the same process with respect to all the frequency points, in which the unit circle is divided, in order to evaluate the phase spectrum of the signal.

As for the 1\textsuperscript{st} method of implementation of the ‘whitened’ Hartley spectrum via the z-transform (algorithm 2), for the 2\textsuperscript{nd} method of implementation of the ‘whitened’ Hartley spectrum via the z-transform, the procedure is based on equation (5.6A) and it can be described as follows.

Initially, the first ‘zero’ is considered, with respect to a frequency point \( w \) (equation (5.7A)), i.e.

\[
z_{w_k} = (\Re(z_k) -\cos(w)) + j(\Im(z_k) -\sin(w)) \tag{5.7A}
\]

Then, the next ‘zero’ is considered and is multiplied with the first ‘zero’. Note that, the imaginary operator, \((j)\), is preserved throughout the process, in order to discriminate between the real and the imaginary components. The same process is repeated for all the ‘zeros’ of the signal, with respect to the frequency point \( w \). Consequently, the product of the ‘zeros’, \( z_{wT} \) (equation (5.8A)), with respect to the frequency point \( w \), is a complex number.

\[
z_{wT} = \prod_{k=1}^{M} z_{w_k} \tag{5.8A}
\]

Subsequently, based on equation (5.6A), the imaginary component of \( z_{wT} \) is subtracted from the real component and then, is divided by its absolute value, thus:

\[
\varphi_{wT} = \frac{\Re(z_{wT}) -\Im(z_{wT})}{\sqrt{\Re^2(z_{wT}) + \Im^2(z_{wT})}} \tag{5.9A}
\]
Consequently, the phase of the signal with respect to a single frequency point, \( \omega \), is calculated. The same process is repeated ('outer' loop) for each frequency point that the unit circle is divided and hence, the 'whitened' Hartley spectrum via the z-transform \( (2^{nd} \) method), is formed.

\[
Y_{TOTAL} = [\phi_{1T}, \phi_{2T}, ..., \phi_{wT}, ..., \phi_{NT}] \quad (5.10A).
\]

Thus, although both methods of implementation provide exactly the same result, their algorithmic implementation is different.

**Removal of the 'intrinsic' discontinuities (optional)**

Step 5: similar to Step 5 of 'whitened' Hartley spectrum via the z-transform \( (1^{st}) \) method.

**Algorithm 4 - Fourier phase spectrum via the Discrete-Time Fourier Transform (DTFT)**

Step 1: same as Step 1 of Fourier phase spectrum via the z-transform.

Step 2: evaluate the FFT (Fast-Fourier Transform), with frequency length equal to \( N \).

Step 3: evaluate the Fourier phase, \( \phi_{w} \), by calculating the inverse tangent of the imaginary over the real part of the \( w^{th} \) frequency point of the Fourier complex spectrum (equation (5.1), chapter five), i.e.

\[
\phi_{w} = \arctan \left( \frac{\mathbb{I}(S_{w})}{\mathbb{R}(S_{w})} \right),
\]

where, \( \mathbb{I}(S_{w}) \) and \( \mathbb{R}(S_{w}) \) are the imaginary and the real component of the \( w^{th} \), Fourier spectral point, respectively.

Thus, in order to evaluate the phase spectrum of the signal, the aforementioned process is repeated for each frequency point, \( \omega \), of the Fourier complex spectrum and consequently,

\[
\phi_{TOTAL} = [\phi_{1}, \phi_{2}, ..., \phi_{w}, ..., \phi_{N}].
\]

Step 4: apply the 'unwrapping' algorithm so as to compensate the 'extrinsic' discontinuities [41] (subsection 5.1.1, chapter five).
Compensation of the 'intrinsic' discontinuities (optional)

Step 5: Starting from the beginning of the phase spectrum, $\pi$ is added to the rest of the spectrum when the point before the 'critical' one has higher value and vice versa (subsection 5.1.2, chapter five).

Algorithm 5 - 'Whitened' Hartley (Hartley phase) spectrum via the Discrete-Time Hartley Transform (DTHT)

Step 1: same as Step 1 of Fourier phase spectrum via the Discrete-Time Fourier Transform (DTFT).
Step 2: same as Step 2 of Fourier phase spectrum via the Discrete-Time Fourier Transform (DTFT).
Step 3: same as Step 3 of Fourier phase spectrum via the Discrete-Time Fourier Transform (DTFT) up to the stage where $\varphi_n$, is obtained. Subsequently, based on equation (5.12) (chapter five) that establishes the relationship between the 'whitened' Hartley spectrum and the Fourier phase spectrum, the 'whitened' Hartley spectrum equals to:

$$Y_{TOTAL} = [\sin(\varphi_1) + \cos(\varphi_1), \sin(\varphi_2) + \cos(\varphi_2), \ldots, \sin(\varphi_N) + \cos(\varphi_N)]$$

where $\varphi_1, \varphi_2, \ldots, \varphi_N$ are the Fourier phase values.

Compensation of the 'intrinsic' discontinuities (optional)

Step 4: Starting from the beginning of the 'whitened' Hartley spectrum, wherever a 'critical' point is detected, the remainder of this spectrum is then multiplied by $-1$ (subsection 5.3.1, chapter five). Hence, after this scan has taken place, the 'whitened' Hartley spectrum via the DTHT ('preliminary' method), is obtained.
Chapter 5. Phase analysis and preliminary results

Additional scan (‘enhanced’ method)

As will be shown from examples that are provided in chapter six, the ‘whitened’ Hartley spectrum obtained from step 4, still has ‘intrinsic’ discontinuities. Hence, the aim of the additional scan (‘enhanced’ method) is to compensate as many of the remaining discontinuities from the ‘whitened’ Hartley spectrum via the DTHT, obtained from step 4, as possible.

Step 5: The first stage of the ‘enhanced’ method is to construct a new ‘whitened’ Hartley spectrum based on the ‘whitened’ Hartley spectrum, obtained from step 4. From now on, this new signal will be referred to as the ‘artificial’ ‘whitened’ Hartley spectrum.

Analytically, the ‘artificial’ ‘whitened’ Hartley spectrum is constructed in the following way:

as known (equation (5.12), chapter five), the general formula of the ‘whitened’ Hartley spectrum is:

\[ Y(\omega) = \sin(\varphi(\omega)) + \cos(\varphi(\omega)) = \sqrt{2} \sin(\varphi(\omega) + \frac{\pi}{4}) \]  

The same equation can be rewritten as:

\[ Y'(\omega) = \sqrt{2} \sin(k\omega + \frac{\pi}{4}) \]

Hence, the ‘artificial’ ‘whitened’ Hartley spectrum, \( Y'(\omega) \), is a sinusoidal signal with amplitude equal to \( \sqrt{2} \), shifted by \( \frac{\pi}{4} \) with respect to the origin and its frequency equals:

\[ 2\pi \frac{\zeta}{2\mu} \]  

where \( \zeta \) and \( \mu \) are the number of zero-crossings and the length of the ‘whitened’ Hartley spectrum, obtained from step 4, respectively.
Chapter 5. Phase analysis and preliminary results

Note that, the frequency, of the 'artificial' 'whitened' Hartley spectrum, \( Y'(\omega) \), does not necessarily have to be exactly equal to the frequency of the 'whitened' Hartley spectrum obtained from step 4. This occurs, because as will be shown from examples provided in chapter six, the 'artificial' 'whitened' Hartley spectrum is only used as a 'reference' signal (i.e. in a similar way as in modulation/demodulation of communication systems).

Then, the 'artificial' 'whitened' Hartley spectrum, \( Y'(\omega) \), is compared with the 'whitened' Hartley spectrum, obtained from step 4. The comparison between the two signals is obtained by subtracting the 'whitened' Hartley spectrum, obtained from step 4, from the 'artificial' 'whitened' Hartley spectrum. As will be shown in examples provided in chapter six (subsection 6.1.4), the majority of the values resulting from the difference between the two spectra are bounded within a certain range. However, a small minority of them are located outside this range. The locations, where these 'outliers' appear, correspond to the frequencies of these remaining 'intrinsic' discontinuities. Thus, whenever the difference is higher than a certain threshold, the locations of the corresponding frequency points are stored as 'critical' points. Then, the 'whitened' Hartley spectrum, obtained from step 4, is scanned from the beginning. Whenever, one of these currently detected 'critical' points is reached, the rest of the spectrum is multiplied by -1 (same as compensation of 'intrinsic' discontinuities step 4). Hence, after this new scan has taken place, the 'whitened' Hartley spectrum via the DTHT ('enhanced' method), is obtained.

Algorithm 6 - 'Whitened' Fourier spectrum

Step 1: same as Step 1 of Fourier phase spectrum via the Discrete-Time Fourier Transform (DTFT).
Step 2: same as Step 2 of Fourier phase spectrum via the Discrete-Time Fourier Transform (DTFT).
Chapter 5. Phase analysis and preliminary results

Step 3: The \( w^{th} \) frequency point of the 'whitened' Fourier spectrum, \( \varphi_w \), is evaluated by dividing the \( w^{th} \) frequency point of the Fourier complex spectrum (\( S_w \)) by its corresponding magnitude, i.e.

\[
\varphi_w = \frac{\mathcal{R}(S_w) + j\mathcal{I}(S_w)}{\sqrt{\mathcal{R}^2(S_w) + \mathcal{I}^2(S_w)}} \quad (5.12A)
\]

where, \( \mathcal{I}(S_w) \) and \( \mathcal{R}(S_w) \) are the imaginary and the real components of the spectrum at the \( w^{th} \) frequency point, respectively.

Thus, in order to evaluate the 'whitened' Fourier spectrum of the signal, the aforementioned process is repeated for each frequency point, \( w \), of the Fourier complex spectrum and consequently,

\[
\varphi_{TOTAL} = [\varphi_1, \varphi_2, \ldots, \varphi_w, \ldots, \varphi_N].
\]

Note that, the 'whitened' Fourier spectrum is complex, it does not convey 'extrinsic' discontinuities (no use of the inverse tangent function) however, there is not a straightforward method that can be used in order to overcome its 'intrinsic' discontinuities. Details related to the advantages and the disadvantages of the 'whitened' Fourier spectrum will be provided in chapter six.

5.6 Summary and conclusions

In this chapter, the implementation of the phase spectrum based on the Fourier (section 5.1) and on the Hartley (section 5.3) transforms, were presented. Initially, the sources of the discontinuities of the conventional Fourier phase function were described, as well as the methods to overcome them. The first category of the discontinuities ('extrinsic') is related to the use of the inverse tangent function and is overcome using the 'unwrapping' algorithm (subsection 5.1.1). The second category of discontinuities ('intrinsic')
originates from the properties of the signal itself and is overcome either with their compensation (subsection 5.1.2) or their removal (subsection 5.1.3) from the phase spectrum. As mentioned, the 'extrinsic' discontinuities exist only in the phase spectrum implementation via the DTFT whereas, the 'intrinsic' discontinuities exist in all the cases (i.e., via the DTFT, via the DTHT, via the z-transform). Examples of synthetic signals were provided, so as to demonstrate the application of the phase cepstrum to time localization (section 5.2).

The major disadvantage of the phase implementation via the DTFT, is the existence of the 'extrinsic' discontinuities in the spectrum (subsection 5.1.1). The 'unwrapping' ambiguities do not allow the application of the Fourier phase cepstrum for the time localization of more than a single pulse as well as the time localization of a pulse in a noisy environment (chapter six).

The 'whitened' Hartley spectrum does not use the inverse tangent function and consequently, it does not have 'wrapping' ambiguities (section 5.3). Hence, the lack of the 'wrapping' ambiguities, allows the Hartley phase cepstrum to localize more than a single pulse (chapter six). The 'whitened' Fourier spectrum also does not use the inverse tangent function and consequently, no 'wrapping' ambiguities arise. However, there is not a known method in order to overcome its 'intrinsic' discontinuities.

In the sixth and the seventh chapters, the Fourier phase and the Hartley phase spectra are used for the time localization of a signal and the audio classification applications, respectively. Hence, in the sixth chapter, the comparison of the Fourier phase cepstrum (section 5.2) with the Hartley phase cepstrum (section 5.4) is extended for the case where the environment is noisy (synthetic signals). In the seventh chapter features are extracted from the Fourier phase and the Hartley phase spectra for the application of audio classification (non-synthetic signals). The aim of both applications is to demonstrate that the phase function conveys meaningful information that is useful for the time localization of a signal and the audio classification applications. Moreover, as will be shown (chapters six and seven), both applications indicate that the phase content of a signal is
encapsulated more efficiently in the Hartley phase rather than in the Fourier phase function.
CHAPTER 6
EXPERIMENTAL RESULTS OF SYNTHETIC SIGNALS
FOR THE TIME LOCALISATION APPLICATION

Introduction

The current (chapter six) and the following (chapter seven) chapter present two applications based on the theory of chapter five. The first application (chapter six) is the use of the Fourier / Hartley phase cepstrum as a time localization tool in a noisy environment.

Hence, in this chapter the comparison of the Fourier phase cepstrum with the Hartley phase cepstrum ('whitened' Hartley cepstrum) (chapter five) is extended for the case where there is noise in the time domain. As with chapter five, the signals used are synthetic so as every step of the process to be easily followed. Examples that demonstrate the superiority of the 'whitened' Hartley spectrum as compared to its Fourier counterpart for the time localization of a signal application in a noisy environment are provided in subsections 6.1.1 to 6.1.3. Moreover, in subsection 6.1.4, an extended version of the compensation rule ('enhanced' method) applied for the reduction of the 'intrinsic'
discontinuities of the ‘whitened’ Hartley spectrum (step 5 - algorithms 5 – section 5.5 – chapter five), is presented via examples. Furthermore, in section 6.2 the behavior of the ‘whitened’ Hartley spectrum in a ‘simulated’ scenario is tested. In section 6.3, the probability density function of the ‘whitened’ Hartley spectrum (Hartley phase spectrum) in the case where noise is added in the time domain, is presented. Finally, section 6.4 summarizes the chapter.

6.1 Experimental results of synthetic signals

In this section, the investigation into the time localization of a signal problem is extended to the case where there is noise added to the time domain. The advantages of the Hartley phase function, over its Fourier counterpart, will be demonstrated. The superiority of the Hartley phase function will be shown via the application to synthetic signals. The method of phase extraction is based on the analysis of the phase cepstrum. Details of the phase cepstral function are given in sections 5.2 and 5.4 of chapter five.

The signals tested here are limited to the rectangular and the exponentially decaying sine wave pulses. The rectangular pulse is chosen, as its shape is very popular amongst communication related applications whereas, the exponentially decaying sine wave pulse is the typical output of a minimum phase system, e.g. human vocal tract. The choice of the signals here is arbitrary, but as already mentioned, the rectangular pulse is characterized by ‘zeros’ located on the unit circle while, the exponentially decaying sine wave pulse is characterized by a set of ‘zeros’ lying inside the unit circle (figures 5.6 and 5.7, chapter five). Thus, these two signals are suitable tools for investigating the issue of time localization.

The phase spectra used for the time localization of a signal application are: the Fourier phase spectrum via the DTFT with/without the compensation of the ‘intrinsic’ discontinuities (algorithm 4 – section 5.5 – chapter five), the ‘whitened’ Fourier spectrum (section 5.4 of chapter five and algorithm 6 – section 5.5 – chapter five) and the ‘whitened’ Hartley spectrum via the DTHT with/without the compensation of the ‘intrinsic’ discontinuities (algorithm 5 – section 5.5 – chapter five). Neither methods of
phase spectrum implementation via the z-transform (i.e. the Fourier phase spectrum via the z-transform (algorithm 1 – section 5.5 – chapter five) and the ‘whitened’ Hartley spectrum via the z-transform (algorithm 2 – section 5.5 – chapter five)) are used due to the errors that the ‘roots’ algorithm produces.

The remaining part of this section is divided into four subsections. In subsections 6.1.1, 6.1.2 and 6.1.3, the behavior of the ‘whitened’ Hartley spectrum is compared with the behavior of the Fourier phase spectrum, in a noisy environment. Then, in subsection 6.1.4, the ‘enhanced’ version of the ‘whitened’ Hartley spectrum via the DTHT (algorithm 5 – section 5.5 – chapter five), which is particularly useful for applications where synthetic signals are used, is presented. As a reminder, the experimental results that show how the Fourier and the Hartley phase functions encapsulate the phase related content of a pulse shifted across time, in a noise-free environment, have been presented in chapter five, sections 5.2 and 5.4.

6.1.1 Introduction to Fourier and Hartley phase cepstrums of pulses with noise

For the comparison of the Fourier-based phase spectra with the Hartley-based phase spectra, in terms of noise tolerance, two pulses of different shape (rectangular and exponentially decaying sine wave) are tested. The behavior of these pulses for three different levels of SNR is investigated. Moreover, the ability of the Fourier-based phase spectra and the Hartley-based phase spectra to detect more than a single pulse in a noisy-free/noisy environment is tested.

The signal to noise ratio (SNR) is defined as:

$$SNR = 10\log_{10} \frac{\sum_{i=1}^{N} s_i^2}{\sum_{i=1}^{N} n_i^2} \quad (6.1)$$

where $s_i$ represents samples of the signal (length equal to $N$) and $n_i$ represents samples of the noise (length equal to $N$) added to the signal. The noise signal is a vector with random entries selected from a normal distribution.
Chapter 6. Experimental results for the time localisation application

Note that the aim of the aforementioned simulations is to compare the behavior of the Fourier phase spectrum, the 'whitened' Fourier spectrum and the 'whitened' Hartley spectrum, in the case where noise is added to the signal in the time domain.

For the Fourier-based phase cepstrums, the results are summarized in tables 6.1 to 6.4 whereas, for the Hartley-based phase cepstrums, the results are summarized in tables 6.5 and 6.6. As a reminder, the phase content of a signal in the cepstral domain, is encapsulated in either the amplitude of the 0th cepstral value or the location of the maximum amplitude of the cepstral peak(s) (sections 5.2 and 5.4 of chapter five). For the phase spectra that convey 'extrinsic' discontinuities (algorithm 4 - section 5.5 - chapter five i.e. Fourier phase spectrum via the DTFT with/without the compensation of the 'intrinsic' discontinuities), the phase content of the signal in the corresponding cepstral domain, is encapsulated in the amplitude of the 0th cepstral value whereas, for the phase spectra that do not convey 'extrinsic' discontinuities (algorithms 5 and 6 - section 5.5 - chapter five, i.e. the 'whitened' Hartley spectrum via the DTHT with/without the compensation of the 'intrinsic' discontinuities and the 'whitened' Fourier spectrum) the phase content of the signal in the corresponding cepstral domain, is encapsulated in the location of the maximum amplitude of the cepstral peak(s), rather than the amplitude of the 0th cepstral value.

Hence, table 6.1 (Fourier phase cepstrum via the DTFT) presents the amplitude of the 0th cepstral value for the exponentially decaying sine wave pulse when shifted in time, in the case where the SNR is 8.9 dBs, -1.1 dBs and -11.1 dBs. Similarly, table 6.2 (Fourier phase cepstrum via the DTFT) presents the amplitude of the 0th cepstral value for the rectangular pulse when shifted in time, in the case where the SNR is 20.2 dBs, 10.2 dBs and 0.2 dBs. The choice of the SNRs, for both pulses, was dictated by the change in amplitude of noise in the time domain. Table 6.3 ('whitened' Fourier cepstrum) presents the location of the maximum amplitude of the cepstral peak for the exponentially decaying sine wave pulse when shifted in time, in the case where the SNR is 8.9 dBs, -1.1 dBs and -11.1 dBs. Table 6.4 ('whitened' Fourier cepstrum) presents the location of the maximum amplitude of the cepstral peaks for the rectangular pulse when shifted in time, in the case where the SNR is 20.2 dBs, 10.2 dBs and 0.2 dBs. Table 6.5 ('whitened' Fourier cepstrum) presents the location of the maximum amplitude of the cepstral peak(s), rather than the amplitude of the 0th cepstral value.
Hartley cepstrum) presents the location of the maximum amplitude of the cepstral peak for the exponentially decaying sine wave pulse when shifted in time, in the case where the SNR is 8.9 dBs, -1.1 dBs and -11.1 dBs. Table 6.6 (‘whitened’ Hartley cepstrum) presents the location of the maximum amplitude of the cepstral peak(s) for the rectangular pulse when shifted in time, in the case where the SNR is 20.2 dBs, 10.2 dBs and 0.2 dBs.

One can observe that in certain boxes of the aforementioned tables, beside the indication of the amplitude/location of the cepstral peak(s) the word ‘missed’ is added. This word is used in the cases where the amplitude/location of the cepstral peak(s) do not correspond to the location of the signal in the time domain, as presented in the 1st column of all the aforementioned tables.

6.1.2 Fourier phase cepstrums of pulses with noise

Tables 6.1 and 6.2 present the amplitude of the 0th cepstral value in the Fourier phase cepstrum via the DTFT. The first table, corresponds to the behavior of the exponentially decaying sine wave, whereas the second table, corresponds to the behavior of the rectangular pulse, for certain values of SNR. The 2nd, the 4th and the 6th columns present the amplitude of the 0th cepstral value in the Fourier phase cepstrum, with the ‘extrinsic’ discontinuities compensated from the corresponding Fourier phase spectrum, for the cases of the exponentially decaying sine wave (table 6.1) and the rectangular pulse (table 6.2), respectively. The 3rd, the 5th and the 7th columns present the amplitude of the 0th cepstral value in the Fourier phase cepstrum, with both the ‘extrinsic’ and the ‘intrinsic’ discontinuities compensated from the corresponding Fourier phase spectrum, for the cases of the exponentially decaying sine wave (table 6.1) and the rectangular pulse (table 6.2), respectively. Tables 6.3 and 6.4 that are dedicated to the ‘whitened’ Fourier cepstrum (no use of the inverse tangent function), present the location rather than the amplitude of the cepstral peak(s). Thus, table 6.3, presents the location of the maximum amplitude of the peak in the ‘whitened’ Fourier phase cepstrum, for the exponentially decaying sine wave whereas, table 6.4, presents the location of the maximum amplitude
of the peaks in the ‘whitened’ Fourier phase cepstrum, for the rectangular pulse, for certain values of SNR.

For the Fourier case via the DTFT (tables 6.1 and 6.2), for both of the pulses tested (exponentially decaying sine wave pulse and rectangular pulse), one can observe that their behavior is heavily affected by the presence of noise. Likewise the results obtained from the tests carried out in a noise-free environment (section 5.2, chapter five), the Fourier phase cepstral domain with the compensation of the ‘extrinsic’ discontinuities (2nd, 4th and 6th columns) from the corresponding spectrum, behaves similarly to the Fourier phase cepstral domain with the compensation of the ‘extrinsic’ and the compensation of the ‘intrinsic’ discontinuities (3rd, 5th and 7th columns) from the corresponding spectrum, for all three different levels of SNR. Moreover, one can observe that even when the SNR is high, in most of the cases, the amplitude of the 0th cepstral value in the Fourier phase cepstrum (tables 6.1 and 6.2), does not correspond to the location of the pulse in the time domain (1st column). Consequently, the experimental results indicate that the ‘unwrapping’ algorithm is affected even in the cases where low-levels of noise are present.

Another disadvantage of the Fourier phase cepstrums, with/without the compensation of the ‘intrinsic’ discontinuities, is that they cannot provide the location of more than a single pulse. Thus, when more than a single pulse exists in the time domain, the amplitude of the 0th cepstral value in the Fourier phase cepstrum corresponds to a point in the middle of the pulses in the time domain. The aforementioned disadvantage is caused due to the heuristic and non-invertible nature of the ‘unwrapping’ algorithm.

The highest noise immunity, from all three categories of Fourier-based cepstrums, is provided by the ‘whitened’ Fourier cepstrum (tables 6.3 and 6.4). Moreover, unlike the Fourier-based phase cepstrum via the DTFT (tables 6.1 and 6.2), the ‘whitened’ Fourier cepstrum can indicate the starting and the finishing points of more than a single pulse, even for the case where the SNR is low. However, the ‘whitened’ Fourier spectrum, from which the ‘whitened’ Fourier cepstrum is obtained, conveys a major disadvantage. Due to the definition of the ‘whitened’ Fourier spectrum (algorithm 6 - section 5.5 – chapter five), for certain frequency points, the numerator and the denominator can become
simultaneously zero and consequently, an 'intrinsic' discontinuity appears. Hence, the
disadvantage is that, unlike the Fourier phase spectrum via the DTFT (tables 6.1 and 6.2)
that uses the inverse tangent function (i.e. the discontinuities ('intrinsic'/ 'extrinsic') can
be overcome either by adding or subtracting $\pi$ or multiples of $\pi$), there is not a known
procedure in order to compensate the 'intrinsic' discontinuities from the 'whitened'
Fourier spectrum.
<table>
<thead>
<tr>
<th>discrete-time domain</th>
<th>exponential pulse (SNR: 8.9 dB)</th>
<th>exponential pulse (SNR: -1.1 dB)</th>
<th>exponential pulse (SNR: -11.1 dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cepstral domain - 'extrinsic' &amp; 'intrinsic' discontinuities compensated</td>
<td>cepstral domain - 'extrinsic' &amp; 'intrinsic' discontinuities compensated</td>
<td>cepstral domain - 'extrinsic' &amp; 'intrinsic' discontinuities compensated</td>
</tr>
<tr>
<td>51-80</td>
<td>36 (missed)</td>
<td>24 (missed)</td>
<td>14 (missed)</td>
</tr>
<tr>
<td>151-180</td>
<td>40 (missed)</td>
<td>49 (missed)</td>
<td>9 (missed)</td>
</tr>
<tr>
<td>251-280</td>
<td>97 (missed)</td>
<td>79 (missed)</td>
<td>17 (missed)</td>
</tr>
<tr>
<td>351-380</td>
<td>178 (missed)</td>
<td>117 (missed)</td>
<td>29 (missed)</td>
</tr>
<tr>
<td>451-480</td>
<td>149 (missed)</td>
<td>119 (missed)</td>
<td>25 (missed)</td>
</tr>
</tbody>
</table>

Table 6.1: Amplitude of the 0<sup>th</sup> cepstral value in the Fourier phase cepstral domain for the exponentially decaying sine wave pulse
<table>
<thead>
<tr>
<th>discrete-time domain</th>
<th>rectangular pulse (SNR: 20.2 dB)</th>
<th>rectangular pulse (SNR: 10.2 dB)</th>
<th>rectangular pulse (SNR: 0.2 dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cepstral domain - 'extrinsic'</td>
<td>cepstral domain - 'extrinsic' &amp;</td>
<td>cepstral domain - 'extrinsic' &amp;</td>
</tr>
<tr>
<td></td>
<td>discontinuities compensated</td>
<td>'intrinsic' discontinuities</td>
<td>'intrinsic' discontinuities</td>
</tr>
<tr>
<td></td>
<td></td>
<td>compensated</td>
<td>compensated</td>
</tr>
<tr>
<td>51-80</td>
<td>68</td>
<td>54</td>
<td>47 (missed)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>25 (missed)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>15 (missed)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>22 (missed)</td>
</tr>
<tr>
<td>151-180</td>
<td>152</td>
<td>142 (missed)</td>
<td>89 (missed)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>58 (missed)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>31 (missed)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>25 (missed)</td>
</tr>
<tr>
<td>251-280</td>
<td>229 (missed)</td>
<td>203 (missed)</td>
<td>136 (missed)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>93 (missed)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>31 (missed)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20 (missed)</td>
</tr>
<tr>
<td>351-380</td>
<td>307 (missed)</td>
<td>267 (missed)</td>
<td>131 (missed)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>96 (missed)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>22 (missed)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>24 (missed)</td>
</tr>
<tr>
<td>451-480</td>
<td>375 (missed)</td>
<td>283 (missed)</td>
<td>163 (missed)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>114 (missed)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>56 (missed)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>40 (missed)</td>
</tr>
</tbody>
</table>

Table 6.2: Amplitude of the 0th cepstral value in the Fourier phase cepstral domain for the rectangular pulse
Table 6.3: Location of the maximum amplitude in the ‘whitened’ Fourier cepstral domain for the exponentially decaying sine wave pulse
### Table 6.4: Location of the maximum amplitude in the ‘whitened’ Fourier cepstral domain for the rectangular pulse

<table>
<thead>
<tr>
<th>discrete-time domain</th>
<th>rectangular pulse (SNR: 20.2 dB)</th>
<th>rectangular pulse (SNR: 10.2 dB)</th>
<th>rectangular pulse (SNR: 0.2 dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>51-80</td>
<td>51-80</td>
<td>51-80</td>
<td>51-80</td>
</tr>
<tr>
<td>151-180</td>
<td>151-180</td>
<td>151-180</td>
<td>151-180</td>
</tr>
<tr>
<td>251-280</td>
<td>251-280</td>
<td>251-280</td>
<td>251-280</td>
</tr>
<tr>
<td>351-380</td>
<td>351-380</td>
<td>351-380</td>
<td>351-380</td>
</tr>
<tr>
<td>451-480</td>
<td>451-480</td>
<td>451-480</td>
<td>451-480</td>
</tr>
</tbody>
</table>
6.1.3 Hartley phase cepstrums of pulses with noise

Tables 6.5 and 6.6 present the location of the maximum amplitude of the cepstral peak(s), for the Hartley phase cepstrum via the DTHT for the exponentially decaying sine wave (table 6.5) and the rectangular pulse (table 6.6), for certain values of SNR. The 2nd, 4th and the 6th columns of tables 6.5 and 6.6, present the location of the maximum amplitude of the cepstral peak(s), in the Hartley phase cepstrum, with the ‘intrinsic’ discontinuities not compensated from the corresponding Hartley phase spectrum, for the cases of the exponentially decaying sine wave and the rectangular pulse, respectively. The 3rd, the 5th and the 7th columns of tables 6.5 and 6.6, present the location of the maximum amplitude of the cepstral peak, in the Hartley phase cepstrum, with the ‘intrinsic’ discontinuities compensated from the corresponding Hartley phase spectrum, for the cases of the exponentially decaying sine wave and the rectangular pulse, respectively.

As for the Fourier case, in the Hartley case, the lower the SNR, the worst the performance of the signal, in the cepstral domain (tables 6.5 and 6.6). The experimental results, presented in tables 6.5 and 6.6, indicate that the ‘whitened’ Hartley spectrum without the compensation of the ‘intrinsic’ discontinuities (2nd, 4th and 6th columns), is more immune to noise compared to the ‘whitened’ Hartley spectrum with the compensation of the ‘intrinsic’ discontinuities (3rd, 5th and 7th columns), for both pulses (exponentially decaying sine wave and rectangular), for all three different levels of SNR. Hence, figure 6.1a presents a pulse in the time domain, located from sample 151 to sample 180, with a 10.2 dB SNR. One can observe that the corresponding ‘whitened’ Hartley cepstrum without the compensation of the ‘intrinsic’ discontinuities (figure 6.1b), is more immune to noise compared to the ‘whitened’ Hartley cepstrum with the compensation of the ‘intrinsic’ discontinuities (figure 6.1c).
Chapter 6. Experimental results for the time localisation application

Figure 6.1a: Time domain of a rectangular pulse with 10.2 dB SNR
Figure 6.1b: Cepstral domain – ‘intrinsic’ discontinuities not compensated
Figure 6.1c: Cepstral domain – ‘intrinsic’ discontinuities compensated
Table 6.5: Location of the maximum amplitude in the Hartley phase cepstral domain for the exponentially decaying sine wave pulse

<table>
<thead>
<tr>
<th>discrete-time domain</th>
<th>exponential pulse (SNR: 8.9 dB)</th>
<th>exponential pulse (SNR: -1.1 dB)</th>
<th>exponential pulse (SNR: -11.1 dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cepstral domain – no discontinuities compensated</td>
<td>cepstral domain – ‘intrinsic’ discontinuities compensated</td>
<td>cepstral domain – no discontinuities compensated</td>
</tr>
<tr>
<td></td>
<td>51-80 53 53</td>
<td>51-80 54 52</td>
<td>51-80 54 2012 (missed)</td>
</tr>
<tr>
<td></td>
<td>151-180 153 153</td>
<td>151-180 153 159</td>
<td>151-180 153 1830 (missed)</td>
</tr>
<tr>
<td></td>
<td>251-280 253 253</td>
<td>251-280 253 253</td>
<td>251-280 253 176 (missed)</td>
</tr>
<tr>
<td></td>
<td>451-480 453 458</td>
<td>451-480 453 4 (missed)</td>
<td>451-480 454 220 (missed)</td>
</tr>
<tr>
<td>discrete-time domain</td>
<td>rectangular pulse (SNR: 20.2 dB)</td>
<td>rectangular pulse (SNR: 10.2 dB)</td>
<td>rectangular pulse (SNR: 0.2 dB)</td>
</tr>
<tr>
<td>----------------------</td>
<td>----------------------------------</td>
<td>----------------------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td></td>
<td>cepstral domain – no discontinuities compensated</td>
<td>cepstral domain – ‘intrinsic’ discontinuities compensated</td>
<td>cepstral domain – no discontinuities compensated</td>
</tr>
<tr>
<td>51-80</td>
<td>51-80</td>
<td>75</td>
<td>51-80</td>
</tr>
<tr>
<td>151-180</td>
<td>151-180</td>
<td>159</td>
<td>151-181</td>
</tr>
<tr>
<td>251-280</td>
<td>251-280</td>
<td>256</td>
<td>250-251</td>
</tr>
<tr>
<td>351-380</td>
<td>351-380</td>
<td>353</td>
<td>350-380</td>
</tr>
<tr>
<td>451-480</td>
<td>451-480</td>
<td>474</td>
<td>450-451</td>
</tr>
</tbody>
</table>

Table 6.6: Location of the maximum amplitude in the Hartley phase cepstral domain for the rectangular pulse
Summarizing, as shown from tables 6.1 to 6.4, from all the Fourier-based cepstrums, only the ‘whitened’ Fourier cepstrum can tolerate noise. Moreover, only the ‘whitened’ Fourier cepstrum can indicate the location of more than a single pulse in the time domain, even for the case where noise is present. As presented in tables 6.5 and 6.6, from the Hartley-based cepstrums, the ‘whitened’ Hartley cepstrum without the compensation of the ‘intrinsic’ discontinuities is more immune to noise compared to the ‘whitened’ Hartley cepstrum with the compensation of the ‘intrinsic’ discontinuities. This occurs because the presence of noise makes the compensation more difficult due to the increased number of the ‘intrinsic’ discontinuities. Finally, as will be shown in section 6.2, the ‘whitened’ Hartley cepstrum without the compensation of the ‘intrinsic’ discontinuities, can also indicate the location of more than a single pulse in the time domain, even for the case where noise is present.

Although, the ‘whitened’ Hartley cepstrum without the compensation of the ‘intrinsic’ discontinuities and the ‘whitened’ Fourier cepstrum demonstrate similar noise immunity, the latter has certain disadvantages. Thus, as mentioned in subsection 6.1.2, it is possible for certain frequency points, the numerator and the denominator of the ‘whitened’ Fourier spectrum to become simultaneously zero and consequently, an ‘intrinsic’ discontinuity to arise. However, for certain applications, where it is necessary to compensate the ‘intrinsic’ discontinuities from the spectrum, there is not a known method in order to overcome them for the ‘whitened’ Fourier case.

Concluding, the ‘whitened’ Hartley cepstrum without the compensation of the ‘intrinsic’ discontinuities is appropriate for applications where noise of considerable power is present. However, for applications where the ‘intrinsic’ discontinuities are necessary to be removed/reduced, then compensation rule can be applied it.

6.1.4 ‘Enhanced’ method

In the third set of simulations, two examples of the ‘enhanced’ approach of the ‘whitened’ Hartley spectrum via the DTHT (algorithms 5 – section 5.5 – chapter five), are presented. The ‘enhanced’ method is the fifth step of algorithm 5 (‘whitened’
Chapter 6. Experimental results for the time localisation application

Hartley spectrum via the DTHT) presented in section 5.5 – chapter five. Note that in order to reach the fifth step of the process, initially the ‘whitened’ Hartley spectrum (step 3 - algorithms 5 – section 5.5 – chapter five) and then, the ‘whitened’ Hartley spectrum with the compensation of the ‘intrinsic’ discontinuities (step 4 - algorithms 5 – section 5.5 – chapter five), have to be implemented. Moreover, as mentioned in the last part of subsection 6.1.3, the ‘whitened’ Hartley cepstrum without the compensation of the ‘intrinsic’ discontinuities (step 3 - algorithms 5 – section 5.5 – chapter five) is appropriate for applications where noise of considerable power is present, whereas the ‘whitened’ Hartley cepstrum with the compensation of the ‘intrinsic’ discontinuities (step 4 - algorithms 5 – section 5.5 – chapter five) is appropriate for applications where the ‘intrinsic’ discontinuities have to be removed/reduced.

The ‘whitened’ Hartley (Hartley phase) spectrum conveys only the ‘intrinsic’ category of discontinuities. Hence, in theory, after the fourth step (algorithms 5 – section 5.5 – chapter five) all the ‘intrinsic’ discontinuities should have been compensated from the Hartley phase spectrum. However, the compensation of the ‘intrinsic’ discontinuities (subsection 5.1.2, chapter five) reduces, rather than eliminates, the phase transitions because the domain is discrete and not continuous (due to the limited sampling rate the compensation rule does not always provide the exact correction (figure 5.2b, chapter five). Hence, the aim of the ‘enhanced’ method (step 5 - algorithms 5 – section 5.5 – chapter five) is to correct these ambiguities. The two examples that follow, provide a step by step description of the algorithm, that performs best for synthetic signals, compared to the other five algorithms presented in section 5.5 – chapter five.

The ‘enhanced’ method can be applied to pulses of any shape however, for the first example, a signal consisted of five periodically spaced exponentially decaying sine waves was selected as the test signal whereas, for the second example, a signal consisted of three periodically spaced exponentially decaying sine waves was selected as the test signal. As a reminder, the phase spectrum of a single exponentially decaying sine wave pulse does not convey ‘intrinsic’ discontinuities (section 5.2 of chapter five). However, the phase spectrum of two or more exponentially decaying sine wave pulses
conveys 'intrinsic' discontinuities. This can be verified by evaluating the 'zeros' of the two (or more) exponentially decaying sine wave pulses (i.e. calculate the 'roots' of the polynomial formed from the coefficients of the signal in the discrete-time domain). A certain number of the 'zeros' of this signal are located on the circumference of the unit circle.

The first set of figures (figures 6.2, 6.3, 6.4 and 6.5) presents the 'enhanced' method applied to the five periodically spaced exponentially decaying sine waves. Figures 6.2a, 6.2b and 6.2c, show the representation of the signal in the time domain, the 'whitened' Hartley spectrum via the DTHT without the compensation of the 'intrinsic' discontinuities and the corresponding 'whitened' Hartley cepstral domain, respectively. Figures 6.3a, 6.3b and 6.3c, show the time domain signal representation (same as figure 6.2a), the 'whitened' Hartley spectrum via the DTHT with the 'intrinsic' discontinuities compensated (step 4 - algorithm 5 - section 5.5 - chapter five) and the corresponding 'whitened' Hartley cepstral domain, respectively. One can observe that the 'whitened' Hartley spectrum of figure 6.3b still has discontinuities and consequently, the corresponding 'whitened' Hartley cepstrum (figure 6.3c) still has phase information (cepstral peaks of lower amplitude) distributed across the phase cepstrum domain. The objective of compensating all 'intrinsic' discontinuities is to concentrate the phase information into one specific location on the phase cepstrum axis. Thus, the aim of the 'enhanced' method is to compensate for these remaining discontinuities of the 'whitened' Hartley spectrum. Initially, the 'zero-crossings' of the 'whitened' Hartley spectrum shown in figure 6.3b, with respect to the frequency axis, are counted (i.e., calculation of the frequency of the spectrum). Subsequently, in step 5 of algorithm 5 (section 5.5 - chapter five), a new sinusoidal signal is constructed. This new sinusoidal signal, referred to as the 'artificial' 'whitened' Hartley spectrum, is presented in figure 6.5a. Then, the 'artificial' 'whitened' Hartley spectrum (figure 6.5a) is compared with the 'whitened' Hartley spectrum of figure 6.3b. Figure 6.5b presents the comparison between the two signals, which is obtained by subtracting the 'whitened' Hartley spectrum of figure 6.3b, from the 'artificial' 'whitened' Hartley spectrum of figure 6.5a. Moreover, figure 6.5c presents the derivative (i.e. 'difference', the first point is subtracted from the second point etc. along the x-axis) of figure 6.5b. Thus, from figure 6.5c, one can observe that the
majority of the values are bounded within a certain range. However, a small minority of them (22) is located outside this range. The locations where these ‘outliers’ appear correspond to the frequencies of these remaining ‘intrinsic’ discontinuities in the ‘whitened’ Hartley spectrum (figure 6.3b). Thus, whenever the ‘difference’ between two consecutive points of figure 6.5c is over a certain threshold (for this case 0.01), the locations (indices) of the corresponding outliers are stored as ‘critical’ points. Then, the ‘whitened’ Hartley spectrum of figure 6.3b is scanned from the beginning. Whenever a ‘critical’ point is reached, the rest of the spectrum is multiplied with −1 (same as the compensation of the ‘intrinsic’ discontinuities, (step 4 – algorithm 5 - section 5.5 – chapter five)). Subsequently, after this compensation has taken place, the ‘whitened’ Hartley spectrum of figure 6.4b (‘enhanced’ method) and the corresponding ‘whitened’ Hartley cepstrum (figure 6.4c), are obtained. Hence, one can observe that due to the compensation of these remaining ‘intrinsic’ discontinuities from the ‘whitened’ Hartley spectrum (figure 6.3b), the ‘whitened’ Hartley cepstrum (‘enhanced’ method) of figure 6.4c is localized compared to the ‘whitened’ Hartley cepstrum of figure 6.3c and the ‘whitened’ Hartley cepstrum of figure 6.2c. Finally, figure 6.5d presents the ‘difference’ of the subtraction of the ‘whitened’ Hartley spectrum (‘enhanced’ method) of figure 6.4b, from the ‘artificial’ ‘whitened’ Hartley spectrum (figure 6.5a). As expected, unlike in figure 6.5c, there are no ‘outliers’ present any more.
Chapter 6. Experimental results for the time localisation application

Figure 6.2a: Time domain of five periodically spaced exponentially decaying sine waves
Figure 6.2b: 'Whitened' Hartley spectrum – 'intrinsic' discontinuities not compensated
Figure 6.2c: 'Whitened' Hartley cepstrum
Figure 6.3a: Time domain of five periodically spaced exponentially decaying sine waves
Figure 6.3b: ‘Whitened’ Hartley spectrum – ‘intrinsic’ discontinuities compensated
Figure 6.3c: ‘Whitened’ Hartley cepstrum
Figure 6.4a: Time domain of five periodically spaced exponentially decaying sine waves
Figure 6.4b: ‘Whitened’ Hartley spectrum – ‘intrinsic’ discontinuities compensated - ('enhanced') method
Figure 6.4c: ‘Whitened’ Hartley cepstrum
Chapter 6. Experimental results for the time localisation application

131

Frequency domain - "artificial" whitened Hartley spectrum:

Subtraction of the "whitened" Hartley spectrum (preliminary) from the "artificial" whitened Hartley spectrum

Difference of the subtraction of the "whitened" Hartley spectrum (preliminary) from the "artificial" whitened Hartley spectrum

Difference of the subtraction of the "whitened" from the "artificial" whitened Hartley spectrum - (figures 6.3b, 6.5a)

Figure 6.5d: Difference of the subtraction of the "whitened" from the "artificial" whitened Hartley spectrum - (figures 6.4b, 6.5a)
Chapter 6. Experimental results for the time localisation application

The second example presents the application of the ‘enhanced’ method to three, instead of five, periodically spaced exponentially decaying sine waves. As with the first example, the first group of the three figures, 6.6a, 6.6b and 6.6c, show the representation of the signal in the time domain, the ‘whitened’ Hartley spectrum via the DTHT without the compensation of the ‘intrinsic’ discontinuities and the corresponding ‘whitened’ Hartley cepstral domain, respectively. The second group of the three figures, 6.7a, 6.7b and 6.7c, show the time domain signal representation (same as figure 6.6a), the ‘whitened’ Hartley spectrum via the DTHT with the ‘intrinsic’ discontinuities compensated (‘preliminary’ method) (step 4 - algorithm 5 - section 5.5 - chapter five) and the corresponding ‘whitened’ Hartley cepstrum domain, respectively. Figure 6.9a presents the ‘artificial’ ‘whitened’ Hartley spectrum which is constructed based on the ‘zero crossings’, with respect to the frequency axis, of the ‘whitened’ Hartley spectrum derived from the ‘preliminary’ method (figure 6.7b). Figure 6.9b shows the result of the subtraction of the ‘whitened’ Hartley spectrum derived from the ‘preliminary’ method (figure 6.7b) from the ‘artificial’ ‘whitened’ Hartley spectrum (figure 6.9a). Figure 6.9c presents the derivative (i.e. ‘difference’, the first point is subtracted from the second point etc. along the x-axis) of figure 6.9b. Similarly to the previous example, from figure 6.9c, one can observe that the majority of the values are bounded within a certain range. However, a small minority of them is located outside this range. The third group of the three figures 6.8a, 6.8b and 6.8c, show the time domain signal representation (same as figure 6.6a), the ‘whitened’ Hartley spectrum via the DTHT with the ‘intrinsic’ discontinuities compensated (‘enhanced’ method) (step 5 - algorithm 5 - section 5.5 - chapter five) and the corresponding ‘whitened’ Hartley cepstrum domain, respectively. Finally, figure 6.9d presents the ‘difference’ of the subtraction of the ‘whitened’ Hartley spectrum (‘enhanced’ method) of figure 6.8b, from the ‘artificial’ ‘whitened’ Hartley spectrum (figure 6.9a). As expected, unlike in figure 6.9c, there are no ‘outliers’ present any more.
Figure 6.6a: Time domain of three periodically spaced exponentially decaying sine waves
Figure 6.6b: ‘Whitened’ Hartley spectrum - ‘intrinsic’ discontinuities not compensated
Figure 6.6c: ‘Whitened’ Hartley cepstrum
Figure 6.7a: Time domain of three periodically spaced exponentially decaying sine waves
Figure 6.7b: ‘Whitened’ Hartley spectrum – ‘intrinsic’ discontinuities compensated
Figure 6.7c: ‘Whitened’ Hartley cepstrum
Figure 6.8a: Time domain of three periodically spaced exponentially decaying sine waves

Figure 6.8b: ‘Whitened’ Hartley spectrum – ‘intrinsic’ discontinuities compensated – (‘enhanced’) method

Figure 6.8c: ‘Whitened’ Hartley cepstrum
Chapter 6. Experimental results for the time localisation application

Figure 6.9a: ‘Artificial’ ‘whitened’ Hartley spectrum

Figure 6.9b: Subtraction of the ‘whitened’ from the ‘artificial’ ‘whitened’ Hartley spectrum

Figure 6.9c: Difference of the subtraction of the ‘whitened’ from the ‘artificial’ ‘whitened’ Hartley spectrum - (figures 6.7b, 6.9a)

Figure 6.9d: Difference of the subtraction of the ‘whitened’ from the ‘artificial’ ‘whitened’ Hartley spectrum - (figures 6.8b, 6.9a)
Concluding, the ‘whitened’ Hartley spectrum via the DTHT without the compensation of the ‘intrinsic’ discontinuities (step 3 – algorithm 5 - section 5.5 – chapter five), the ‘whitened’ Hartley spectrum with the compensation of the ‘intrinsic’ discontinuities (‘preliminary’ method) (step 4 – algorithm 5 - section 5.5 – chapter five) and the ‘whitened’ Hartley spectrum with the compensation of the ‘intrinsic’ discontinuities (‘enhanced’ method) (step 5 – algorithm 5 - section 5.5 – chapter five), form the three evolutionary steps of the same process. The advantages and the disadvantages of these three Hartley-based phase spectra can be summarized as follows.

The ‘whitened’ Hartley spectrum without the compensation of the ‘intrinsic’ discontinuities (step 3 – algorithm 5 - section 5.5 – chapter five) demonstrates the highest noise immunity compared to the other two and hence, it can detect the location of pulse(s) even when noise of high power is present. Moreover, the ‘whitened’ Hartley spectrum without the compensation of the ‘intrinsic’ discontinuities is appropriate for the cases where the starting and finishing point of the pulse(s) is required.

The ‘whitened’ Hartley spectrum with the compensation of the ‘intrinsic’ discontinuities (‘preliminary’ method) (step 4 – algorithm 5 - section 5.5 – chapter five) is recommended for applications where an estimation of the middle point of a pulse/group of pulses, is required (figures 6.3 and 6.7). Moreover, the ‘whitened’ Hartley cepstrum with the compensation of the ‘intrinsic’ discontinuities (‘preliminary’ method) is smoother compared to the ‘whitened’ Hartley cepstrum without the compensation of the ‘intrinsic’ discontinuities. However, the ‘whitened’ Hartley spectrum with the compensation of the ‘intrinsic’ discontinuities (‘preliminary’ method) demonstrates lower noise immunity compared to the ‘whitened’ Hartley spectrum without the compensation of the ‘intrinsic’ discontinuities.

Finally, the ‘whitened’ Hartley spectrum with the compensation of the ‘intrinsic’ discontinuities (‘enhanced’ method) (step 5 – algorithm 5 - section 5.5 – chapter five) conveys the least amount of ‘intrinsic’ discontinuities and hence, it produces the smoothest cepstrum (figures 6.4 and 6.8), compared to both the aforementioned cases.
However, it demonstrates the lowest noise immunity and thus, it is appropriate in the cases where no noise or noise of low power is present.

6.2 Behavior of the ‘whitened’ Hartley spectrum in a ‘simulated’ scenario

For the fourth set of experiments, the behavior of the ‘whitened’ Hartley spectrum in a ‘simulated’ scenario, is tested. In the example described here, a rectangular pulse is transmitted from a source (e.g., radar) and multiple reflections of it, are received with time delays and varying amplitudes, in a noisy environment (16 dB SNR).

Figure 6.10a shows six ‘snapshots’ of the pulse. The first ‘snapshot’ presents the pulse when initially transmitted from the source (time samples: 31 to 60). However, as the pulse shifts in time, its amplitude eventually decreases, although its power (‘area’ of the pulse) does not change (time samples: 180 to 212). The pulse reflects, because of an obstacle (time samples: 430 to 497) and returns towards the source with its amplitude reducing even further (time samples: 452 to 410, 349 to 300 and 114 to 55) (figure 6.10a). The limits of the accuracy of the time delay estimation, depends on the width of the available time domain window, over which analysis can be made, together with the power of noise added to the signal. Note that, in order to reduce the ambiguities due to aliasing, the signal is shifted only between the 1st to the 512th out of the 2048 points of the time domain.
Chapter 6. Experimental results for the time localisation application

The phase content of the signal is chosen to be presented via the 'whitened' Hartley spectrum without the compensation of the 'intrinsic' discontinuities (step 3 - algorithm 5 - section 5.5 - chapter five) because, as shown in subsection 6.1.3, it demonstrates the highest noise immunity and also is appropriate for the cases where the starting and finishing point of the pulse(s) is required. Thus, one can observe that the ‘whitened’ Hartley cepstrum (figure 6.10b) of the time signal presented in figure 6.10a, encapsulates its phase content accurately. As expected, the starting and finishing points of the pulse/reflection of the pulse in the time domain (figure 6.10a), coincide with the highest cepstral peaks of the ‘whitened’ Hartley cepstrum (figure 6.10b).

Figure 6.10a: Time domain of the ‘simulated’ scenario
Figure 6.10b: Cepstral domain of the ‘simulated’ scenario
6.3 Behavior of the Fourier and the Hartley phase functions in a noisy environment

In subsections 6.1.1 to 6.1.3, the Fourier phase spectrum and the Hartley phase spectrum were compared (via their corresponding cepstrums), with respect to their tolerance in noise. The aim of this section is to explain the reason that the Hartley phase spectrum is more immune to noise when compared to the Fourier phase spectrum, in the time localization of a signal application (section 6.1). In the following part, the probability density function (pdf) of the signal in the Hartley phase and the Fourier phase spectra, in the case where the signal in the time domain is Gaussian noise, will be presented. It is accepted that where the signal in the time domain is not Gaussian noise, the distribution of the samples in the Hartley phase and the Fourier phase spectra may also change.

Assume that the signal in the time domain is Gaussian distributed noise. Hence, in the case where the signal in the time domain is Gaussian distributed noise, the distribution of the same signal, in the Hartley phase spectrum is given by:

\[
p_H(\beta) = \frac{1}{\pi \sqrt{2n}} \frac{1}{\sqrt{1 - \frac{\beta^2}{\sqrt{2}}}}, \quad -\sqrt{2} < \beta < \sqrt{2}
\]

(6.2)

where, \( \beta \) are the Hartley's phase function amplitude values.

The derivation of the probability density function, \( p_H(\beta) \), is presented in appendix A2.

The shape of \( p_H(\beta) \) is shown in figure 6.11a. One can observe from figure 6.11a that, the regions of the pdf where the majority of the samples are concentrated, are in its upper and lower range (i.e. around \( \sqrt{2} \) and \(-\sqrt{2} \)). However, the information content of the signal, in the Hartley phase spectrum, is encapsulated in the zero crossings with respect to the frequency axis (figures 6.2b, 6.2c, 6.3b, 6.3c, 6.4b and 6.4c) rather than in the minimum/maximum values of the cosinusoidal signal (i.e. \( \pm \sqrt{2} \)). Consequently, a noisy signal mainly affects the higher and the lower domain of the Hartley phase spectrum and
hence, its information content (encapsulated in the zero-crossings, middle part of its domain) is less affected.

For the Fourier phase spectrum case though, assuming again that the distribution of the signal in the time domain is Gaussian then, its distribution in the conventional Fourier phase spectrum, is given by:

$$p_F(\beta) = \frac{1}{\pi \sqrt{1 + \beta^2}} \quad (6.3)$$

The probability density function presented in equation (6.3), $p_F(\beta)$, is called Cauchy [49]. The Cauchy distribution, (figure 6.11b) $p_F(\beta)$, similarly to the Gaussian distribution, is symmetrical about $\beta = 0$ with its maximum value at $\beta = 0$. However, the Cauchy distribution falls more rapidly as $|\beta|$ increases and also its tails are heavier, compared to the Gaussian distribution. Nevertheless, for the case of the conventional Fourier phase spectrum, unlike the case of the Hartley phase spectrum, the information content is distributed across the whole range of samples.
Chapter 6. Experimental results for the time localisation application

Figure 6.11a: Probability density function (pdf) of noise in the Hartley phase spectrum
Figure 6.11b: Probability density function (pdf) of noise in the Fourier phase spectrum

Note that, the audio signals (e.g. speech, mechanical sounds etc.), although unlikely to be true Gaussian noise, convey a heavy noise (figure 5.4 of chapter five) component and hence, the pdfs of their phase spectra are similar to the Cauchy (figure 6.11b) - Fourier case and the distribution described from equation (6.2) (figure 6.11a) - Hartley case.

The aforementioned behavior of the Hartley phase function in a noisy environment, is useful for the applications of coding [63] and time localization of a signal (sections 6.1 and 6.2).
6.4 Summary and conclusions

This chapter presents the use of the ‘whitened’ Hartley spectrum (Hartley phase function) as a tool for time delay estimation. This task demonstrates the advantages of the Hartley phase function via its application to synthetic signals. The benefit of using synthetic signals is that it is easier to observe certain details of the process.

Hence, this chapter compares the Hartley phase cepstrum with its Fourier counterpart. The Hartley phase cepstrum (subsection 6.1.3), unlike its Fourier counterpart (subsection 6.1.2), is more tolerant to noise and also, it experiences only the ‘intrinsic’ category of discontinuities, which can be detected and compensated. Moreover, the Fourier phase cepstrum can identify only a single pulse, due to the ambiguities introduced by the use of the ‘unwrapping’ algorithm. However, as shown in subsections 6.1.3 and section 6.2, the ‘whitened’ Hartley cepstrum without the compensation of the ‘intrinsic’ discontinuities, can indicate the location of more than a single pulse in the time domain, even for the case where noise is present.

In section 6.3, the noise immunity of the Hartley phase spectrum, compared to its Fourier counterpart, is justified via the shape of the probability density function of the signal in the Hartley phase spectrum, in the case where the signal in the time domain is Gaussian noise.

Summarizing, the major disadvantage of the implementation of the phase spectrum based on the Fourier (subsection 6.1.2), is the heuristic approach of the compensation of the ‘extrinsic’ discontinuities (‘wrapping’ ambiguities). The effect of the ‘wrapping’ ambiguities is more severe in the case where noise is present (subsection 6.1.2). The advantages of the Hartley phase spectrum, is that it does not convey ‘extrinsic’ discontinuities and also, due to its structure, it is less affected (figure 6.11a) by the presence of noise.
Consequently, this chapter justifies that the Hartley phase function can be used as a substitute to the Fourier phase function, as it encapsulates the phase content of the signal in a more efficient and easy to identify manner.
CHAPTER 7
EXPERIMENTAL RESULTS OF NON-SYNTHETIC SIGNALS FOR THE AUDIO CLASSIFICATION APPLICATION

Introduction

In this chapter, unlike chapter six, the signals involved are non-synthetic. The feature vectors formed from the statistical features extracted from the phase related spectrograms of the audio recordings are used together with the feature vectors formed from the statistical features extracted from the magnitude spectrograms of the audio recordings for the audio classification application.

As mentioned in the literature review (chapter two, sections 2.1 and 2.2), recent research on automatic content-based classification of audio recordings employs the magnitude spectrum and temporal related features in order to classify different acoustic events retrieved from audiovisual databases. In most cases, the task is to attempt to classify sounds that, although belonging to the same family, such as sports sounds, are
acoustically dissimilar. Also, there are other cases in which it is attempted to discriminate between various TV events such as news, advertisements etc. that are again acoustically dissimilar. Some of the features employed for speech recognition, such as Mel-scale cepstral coefficients or Perceptual Linear Prediction coefficients, are not always effective for audio classification, as speech forms a very specialized group of the audio family. Magnitude spectrum features convey important energy-related information that is essential for the classification process. Hence, most of the frequency domain feature extraction techniques, for audio classification, are based on the spectral magnitude content of the signal [95]. However, the calculation of the magnitude spectrum of a signal preserves the information that is related to the absolute value of its real and imaginary Fourier components but it does not preserve the information related to the signs / change of signs of its real and imaginary Fourier spectrum components. The aim of this work is to show that, in the case where the feature vectors formed from the phase related spectrograms of the signals are used, together with the feature vectors formed from the magnitude spectrograms of the signals, provide higher classification score compared to the case where the feature vectors used are formed only from the magnitude spectrograms of the signals.

Hence, the novelty of this work is the use of phase related information (together with magnitude information), for frequency based statistical feature extraction for audio classification.

Most researchers do not use phase information in the classification process, although it conveys important information, because of the difficulties in processing the discontinuities in the phase spectrum. The discontinuities appearing in the phase related spectra are unwanted features that affect the classification rate. As mentioned in chapter five (section 5.1), there are two kinds of discontinuities that appear in the Fourier phase spectrum, the first one arises from the computation of the phase ('extrinsic' discontinuity), whereas the second is due to the structure of the signal ('intrinsic' discontinuity). However, the experimental results that are presented in section 7.3, show that after removing / compensating the discontinuities from the Fourier phase spectrum (section 5.1,
chapter five) and the Hartley phase spectrum (section 5.3, chapter five), the phase related information, assists the classification process (i.e. increases the classification rate), significantly.

Hence, the proposed method is briefly described as follows: each audio recording of the database used (section 7.1), is divided into equal-length frames (256 samples each) and transformed into the frequency domain. Subsequently, a matrix is formed. The first row of the matrix corresponds to the spectrum of the first time frame while, the last row of the matrix, to the spectrum of the last time frame. Note that, in the cases where the length of the last frame of the signal, in the time domain, is less than 256 samples, the signal is zero-padded. Hence, based on the aforementioned process the following spectrograms (time – frequency signal representations), are implemented:

- the Fourier magnitude spectrogram, the Hartley magnitude spectrogram (equation (3.9), subsection 3.1.1, chapter three), the Hartley transform spectrogram (equation (3.1), subsection 3.1.1, chapter three). Moreover, from the phase spectra described in section 5.5 – chapter five, the ones formed as spectrograms in order to be tested for the classification application are: the Fourier phase spectrum via the DTFT with/without the compensation of the ‘intrinsic’ discontinuities (algorithm 4 – section 5.5 – chapter five), the Hartley phase spectrum via the DTHT with/without the compensation of the ‘intrinsic’ discontinuities (algorithm 5 – section 5.5 – chapter five), the Fourier phase spectrum via the z-transform (algorithm 1 – section 5.5 – chapter five) and the Hartley phase spectrum via the z-transform (algorithms 2 – section 5.5 – chapter five). The ‘whitened’ Fourier transform (algorithm 6 – section 5.5 – chapter five) is not tested, because its spectrum is complex and thus, there is not a straightforward method in order to extract features from it. Similarly, the ‘enhanced’ method of the Hartley phase spectrum via the DTHT (step 5 – algorithm 5 - section 5.5 – chapter five) is also not tested, because as mentioned in subsection 6.1.4 (chapter six), its use is recommended for applications where the signals are synthetic and the noise is low. Finally, based on the aforementioned method (i.e. divide each audio recording in frames of 256 samples etc.), the Wigner-Ville distribution (equation (3.38), section 3.4, chapter three) is also implemented.
Chapter 7. Experimental results for the audio classification application

Each of the eight signal representations has to be presented to the classifier with its dimensionality reduced (section 4.4, chapter four). Hence, eight statistical features are calculated from each signal representation for each audio recording. The statistical features calculated are: the variance, the skewness, the kurtosis, the entropy, the inter-quartile range, the range, the median and the mean absolute deviation (section 4.4, chapter four). The statistical values derived from each of the signal representations, form the feature vectors that are independently passed to the Mahalanobis classifier (section 4.6, chapter four).

As will be discussed in section 7.2, the signal representations implemented (i.e. the Fourier magnitude, the Hartley phase via the z-transform, the Hartley phase via the DTHT, the Fourier phase via the z-transform, the Fourier phase via the DTFT, the Wigner – Ville, the Hartley magnitude, the Hartley transform) are compared to each other, in terms of the classification rates they obtain and after selecting the three that perform better, the combinatory classification scheme is employed and the final experimental results (section 7.3) are obtained.

Hence, the remaining part of this chapter is organised as follows. In section 7.1 the database used for the experiments will be described. Section 7.4 is dedicated to four issues that are associated with the phase related spectrograms (i.e. the Hartley phase via the z-transform, the Hartley phase via the DTHT, the Fourier phase via the z-transform and the Fourier phase via the DTFT). It is noted that, to the best of author's knowledge there is no published work related to feature extraction from phase spectrograms and, hence, details related to the sampling rate (subsection 7.4.3), the use of 'difference' (subsection 7.4.1) as well as the effect of the phase function discontinuities in the classification rate (subsection 7.4.2) have to be addressed. Moreover, as mentioned in chapter five, the phase related spectra can be implemented based on either the Fourier transform (algorithms 1 and 4 – section 5.5 – chapter five) or the Hartley transform (algorithms 2 and 5 – section 5.5 – chapter five). Hence, in subsection 7.4.4, there will be a comparison, in terms of classification rate, between the phase spectrograms derived from the Fourier transform with the phase spectrograms derived from the Hartley
transform. In section 7.2, the three signal representations that perform better, in terms of classification rate are selected (subsection 7.2.1) and the decision rule is presented (subsection 7.2.2). Moreover, in the same section, the cross-validation resampling method (subsection 7.2.3) and the significance statistical test (subsection 7.2.4) are also described. The cross-validation resampling method simulates different combinations of test and training vectors so as to reduce the effect of the outliers in the case where the size of the database is small. The significance statistical test indicates whether a classification score is substantially higher compared to another. Finally, in section 7.3, the classification rates of the proposed combinatory scheme are presented while, in section 7.5, the summary of this chapter is given.

7.1 Database

In the experimental part, the database used consists of ten different classes of recordings of gunshots (table 7.1). The audio recordings are 16-bit, mono and their sampling frequency is 44.1 kHz. This database is chosen because all the classes are acoustically similar. Consequently, the discrimination is more difficult when compared to the case where the classes belong acoustically to dissimilar classes.

The length of each audio recording is different, due to the non-artificial nature of the data. Thus, the second column of table 7.1 provides the average length of the recordings, of each class. Each one of the ten classes consists of ten recordings. Seven of them are used as training data and the remaining three as test data.
### Chapter 7. Experimental results for the audio classification application

<table>
<thead>
<tr>
<th>Classes of gunshots</th>
<th>Average length of the recordings of each class (in samples)</th>
</tr>
</thead>
<tbody>
<tr>
<td>i) firing a revolver</td>
<td>35350</td>
</tr>
<tr>
<td>ii) firing a .22 caliber handgun</td>
<td>12360</td>
</tr>
<tr>
<td>iii) firing a M-1 rifle</td>
<td>45230</td>
</tr>
<tr>
<td>iv) firing a World War II German rifle</td>
<td>50740</td>
</tr>
<tr>
<td>v) firing a cannon</td>
<td>63680</td>
</tr>
<tr>
<td>vi) firing a 30-30 rifle</td>
<td>45180</td>
</tr>
<tr>
<td>vii) firing a .38 calibre semi-automatic pistol</td>
<td>12240</td>
</tr>
<tr>
<td>viii) firing a lever action Winchester rifle</td>
<td>33730</td>
</tr>
<tr>
<td>ix) firing a 37mm anti-tank gun</td>
<td>73710</td>
</tr>
<tr>
<td>x) firing a pistol</td>
<td>59700</td>
</tr>
</tbody>
</table>

Table 7.1: Database of gunshots

One common characteristic for all ten classes is the similarity in their Fourier magnitude spectrograms. Specifically, the similarity can be understood from the time evolution of the spectral envelope. Examples of their magnitude spectral shape are shown in figures 7.1 and 7.2.

**Figure 7.1:** Fourier magnitude spectrogram of an audio recording from class (vi)

**Figure 7.2:** Fourier magnitude spectrogram of an audio recording from class (x)
In all the classes of gunshots, a spectral peak, which corresponds to the detonation of the gun, appears at the beginning of the recording. Then, the initial magnitude burst eventually decays.

Another common characteristic between all the classes of gunshots, is their noise content. Noisy signals generate randomly located ‘zeros’ on or very close to the circumference of the unit circle in the z-plane. Figure 5.4 of chapter five shows the location of the ‘zeros’, in the z-plane, for a single frame (256 samples long) taken from one of the recordings that belongs to class (vii). One can observe that the majority of the ‘zeros’ are located on or very close to the circumference of the unit circle, which is an indication of the signal’s noise content. Generally, the ‘zeros’ of non-synthetic signals (such as speech [43]), are similarly located in the z-plane, due to their noise content.

### 7.2 Preliminary classification results, signal representation selection, decision rule, cross-validation and significance test

The process followed in order to obtain the classification scores (table 7.2) for the ten classes of the database (section 7.1), with features extracted from the Fourier magnitude spectrogram, the Hartley magnitude spectrogram, the Hartley transform spectrogram, the Wigner-Ville distribution, the Hartley phase ‘difference’ spectrogram via the DTHT and the Hartley phase ‘difference’ spectrogram via the z-transform, is described as follows.

Initially, the feature vectors, derived from the statistical feature extraction of each signal representation derived from each audio recording, are formed. As mentioned (section 7.1), the database consists of ten classes and each class consists of ten samples. Seven of them are used as training vectors while, the remaining three are used as test vectors. Consequently from the one hundred feature vectors, seventy of them are used as training vectors and the remaining thirty as test vectors, for each signal representation. The mean value of each of the eight statistical measures, from the seven training vectors of each class, forms the ten target vectors (i.e. one for each class).
Then, each one of the test vectors is classified to a certain class based on the minimum distance criterion, i.e. the minimum distance between itself and each one of the target vectors. The same process is repeated for all the thirty test vectors. Hence, the classification score over all the ten classes is obtained from the number of the test vectors classified to the correct class.

Hence, based on the aforementioned process, the classification scores over all the ten classes, for the six signal representations are presented in table 7.2.

<table>
<thead>
<tr>
<th>signal representations</th>
<th>classification score in (%) (over all the 10 classes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>signal representation that conveys magnitude information</td>
<td>Fourier magnitude spectrogram</td>
</tr>
<tr>
<td>signal representations that convey phase information</td>
<td>Hartley phase dif. spectrogram via the z-transform</td>
</tr>
<tr>
<td></td>
<td>Hartley phase dif. spectrogram via the DTHT</td>
</tr>
<tr>
<td>signal representations that convey combined information (i.e. both magnitude and phase)</td>
<td>Hartley magnitude spectrogram</td>
</tr>
<tr>
<td></td>
<td>Hartley transform spectrogram</td>
</tr>
<tr>
<td></td>
<td>Wigner-Ville</td>
</tr>
</tbody>
</table>

Table 7.2: Classification scores in % over all the ten classes
One can observe that from the six signal representations (table 7.2) the highest classification rate is obtained from the Fourier magnitude spectrogram (66.7%). Thus, the experimental results indicate that the feature vectors extracted from the Fourier magnitude spectrograms of the signals tested, encapsulate their content more efficiently compared to the feature vectors extracted from the other signal representations presented in table 7.2. However, the recognition rate, obtained from the Fourier magnitude spectrogram, could be higher in the case where the feature vectors extracted from one or more of the remaining five signal representations would assist the classification process. Hence, the classes that have to be assisted are: the second, the fourth, the fifth and the tenth. These four classes are selected because of the low classification rate they obtain (based on the results derived from the feature vectors extracted from the Fourier magnitude spectrogram) and also, because of the distance between the locations of their feature vectors. In the case where the feature vectors, extracted from the members of the same class, are located far from each other, this indicates high possibility of misclassification when target vectors of other classes are located near [98].

The selection of the signal representations that will participate to the classification process as well as the decision rule used, are described in subsections 7.2.1 and 7.2.2, respectively.

7.2.1 Signal representation selection

As mentioned, the Fourier magnitude spectrogram provides the highest classification rate when compared to the other five signal representations and hence, will be used as the principal feature stream of the combinatory classification scheme (subsection 7.2.2). Moreover, the Fourier magnitude spectrogram is the only one, out of the aforementioned six signal representations, that preserves just the magnitude content of the signal.

Hence, from the remaining five signal representations some of them will be selected so as to form the assistive feature streams of the combinatory classification scheme (subsection 7.2.2). In the following part, the five signal representations will be categorized according
Chapter 7. Experimental results for the audio classification application

to the information they encapsulate, so as to select the ones that convey distinct information and thus, avoid information overlapping;

i) As mentioned, the first category of signal representation(s) includes the Fourier magnitude spectrogram that conveys magnitude spectrum information.

ii) The second category includes the signal representations that preserve only the phase related information of the signal. From this category, a choice had to be made among the Hartley phase ‘difference’ spectrogram via the DTHT and the Hartley phase ‘difference’ spectrogram via the z-transform.

The advantage of the construction of the Hartley phase ‘difference’ spectrogram via the DTHT is that its implementation is based on compensation and thus, there is no need to empirically choose a certain exclusion ‘ring’ width. On the other hand, the advantage of the construction of the Hartley phase ‘difference’ spectrogram via the z-transform is that by varying the width of the exclusion ‘ring’, it is possible to exclude not only the ‘zeros’ that are located on the circumference of the unit circle (‘intrinsic’ discontinuities) but also the ‘sharp zeros’ that cause rapid phase transitions in the phase spectrum as well. Thus, as both approaches have their own advantages both the Hartley phase ‘difference’ spectrogram via the z-transform and the Hartley phase ‘difference’ spectrogram via the DTHT will be used in the combinatory scheme (subsection 7.2.2), one at a time.

iii) Finally, the third category, unlike the aforementioned two categories, includes the signal representations that do not convey purely magnitude or purely phase information. Namely, these signal representations are: the Hartley magnitude spectrogram (based on equation (3.9), subsection 3.1.1, chapter three), the Hartley transform spectrogram (based on equation (3.1), subsection 3.1.1, chapter three) and the Wigner – Ville (based on equation (3.38), section 3.4, chapter three). The experimental results (table 7.2) show that the Hartley
magnitude spectrogram performed better, in terms of classification rate, when compared to either the Hartley transform spectrogram or the Wigner – Ville. Moreover, the feature vectors, extracted from the members of the same class, derived from the Hartley magnitude spectrogram, are located closer to each other when compared to the feature vectors extracted from the members of the same class, derived either from the Wigner-Ville or from the Hartley transform spectrogram [98]. Thus, the Hartley magnitude spectrogram is selected to participate in the combinatory scheme (subsection 7.2.2).

Summarizing, the first group of the signal representations selected to participate in the combinatory scheme is: the Fourier magnitude spectrogram (from category (i)), the Hartley phase ‘difference’ spectrogram via the z-transform (from category (ii)) and the Hartley magnitude spectrogram (from category (iii)). The second group of the signal representations selected to participate in the combinatory scheme is: the Fourier magnitude spectrogram (from category (i)), the Hartley phase ‘difference’ spectrogram via the DTHT (from category (ii)) and the Hartley magnitude spectrogram (from category (iii)).

Note that, the criteria on which the signal representation selection is based or the classes chosen to be assisted (introduction of section 7.2) could be different, however the aim of this chapter is to provide an example, which indicates that phase related information can enhance the recognition rate, when combined with the magnitude information, for the application of frequency-based feature extraction for audio classification, rather than try to find the best combination of signal representations / classes so as to obtain the highest possible classification rate.

7.2.2 Decision rule

As mentioned, the three spectrograms that will participate in the combinatory classification scheme are: the Fourier magnitude (principal feature stream), the Hartley magnitude (assistive feature stream) and the Hartley phase via the z-transform (assistive feature stream), for the 1st experiment and the Fourier magnitude (principal feature
stream), the Hartley magnitude (assistive feature stream) and the Hartley phase via the DTHT (assistive feature stream), for the 2nd experiment.

The decision rule that will be described in the next few lines is a two-stage process. In the first stage, the classification rate will be derived from the feature vectors of the Fourier magnitude spectrogram (principal feature stream) whereas, in the second stage the feature vectors extracted from the two other spectrograms (assistive feature streams) will assist the classification process. Thus, after this two-stage process the final classification rate, is obtained.

**Stage 1**

In the first stage of the classification process, each test vector, derived from the Fourier magnitude spectrogram, is allocated to a certain class based on the minimum distance between itself and each of the ten target vectors (introduction of section 7.2). Thus, depending on the principal feature stream (Fourier magnitude spectrogram), the preliminary decision for the allocation of each one of the thirty test vectors, is obtained.

**Stage 2**

As mentioned in subsection 7.2.1, the feature vectors extracted from the Hartley magnitude spectrogram and the Hartley phase spectrogram via the DTHT will be used as the assistive classification streams. As a reminder (introduction of section 7.2), the aim of the use of the assistive feature streams is to increase the classification scores of the second, the fourth, the fifth and the tenth classes. Preliminary experiments indicated that the feature vectors derived from the Hartley magnitude spectrogram can assist more efficiently, in terms of classification, the classes four and five whereas, the feature vectors derived from the Hartley phase spectrograms the classes two and ten.

Thus, these two assistive feature streams have to assist, in terms of classification, the four classes (2nd, 4th, 5th and 10th). Consequently, instead of measuring the distance of each test
vector with respect to each one of the ten target vectors, as occurs for the principal
classification stream (Fourier magnitude spectrogram, stage 1), it is more useful to
measure the distance of each test vector with respect to the target vector of a class where
the score is to be increased.

Thus, for the first assistive stream (Hartley magnitude spectrogram), the distance between
each one of the thirty test vectors (derived from the Hartley magnitude spectrogram) and
the target vector of class four (again derived from the Hartley magnitude), is measured.
Consequently, the test vectors that are ‘closest’ to the target vector are classified to class
four. The boundary (i.e. the distance that specifies whether a test vector is close or not to
the target vector) is selected based on the following non-heuristic process. The distances
of each test vector from the target vector of class four are organised in ascending order
(i.e. minimum distance, with respect to the target vector, first). Then, the difference (i.e.
the first closest distance minus the second closest distance, the second closest distance
minus the third closest distance etc.) between the sorted distances, is calculated. The
location where the difference of the distances obtains its highest value is the limit where
the boundary is positioned. Hence, the test vectors where their distances to the target
vector are smaller (i.e. located closer) as compared to the distance where the boundary is
located, are classified to class four.

The same process is repeated for the fifth class based again on the Hartley magnitude
spectrogram followed by the second and the tenth classes, both based on the Hartley
phase via the z-transform spectrogram.

Then, the final step is to identify whether the aforementioned test vectors (i.e. the ones
classified based on the Hartley magnitude spectrogram and the Hartley phase via the z-
transform spectrogram) are classified to the same class based on the principal feature
stream (i.e. the Fourier magnitude spectrogram) of stage 1, as well. Hence, in the case
where the principal feature stream ‘disagrees’ with the assistive feature stream for a
certain test vector, then the test vector is classified based on the decision of the assistive
feature stream. After the aforementioned process, for the assistance of the four classes
Chapter 7. Experimental results for the audio classification application

(2\textsuperscript{nd}, 4\textsuperscript{th}, 5\textsuperscript{th} and 10\textsuperscript{th}), has been completed, the final classification rate, over all the ten classes, is obtained, based on the principal feature stream (stage 1).

Thus, as mentioned, for the first experiment, the Fourier magnitude spectrogram is used as the principal feature stream, while the Hartley magnitude and the Hartley phase via the z-transform spectrograms are used as the assistive feature streams.

For the second experiment, the Fourier magnitude spectrogram is used as the principal feature stream, while the Hartley magnitude and the Hartley phase via the DTHT spectrograms are used as the assistive feature streams.

Note that for both the first and the second experiments, in terms of classification score, the sequence in which the assistive feature streams are chosen is unimportant (i.e. the Hartley magnitude first and then the Hartley phase or vice versa) or the sequence in which the classes are selected in order to be assisted. All the possible combinations provide similar classification rates.

The classification scores, obtained from the aforementioned decision rule, for both experiments, are presented in section 7.3.

7.2.3 Cross-validation

As mentioned in section 7.1, the database used has ten classes and each class consists of ten recordings. Hence, each class, of the aforementioned database, consists of a limited number of samples. Thus, a certain number of the possible combinations of the training and the test feature vectors are simulated, so as to reduce the effect caused to the classification rate by the outliers.

Databases that consist of a limited number of samples, is a common problem in the area of pattern recognition and consequently, certain resampling methods [68] have been implemented. In the version of cross-validation used for this research, the samples (i.e.
feature vectors formed from each audio recording) of the database are divided into $l$ numbers of subsets. Each time a test is conducted one of the $l$ subsets is used for testing while, the remaining subsets are used for training. The same process can be repeated until every subset has been used for testing. However, it is not necessary to carry out tests with all the possible combinations. A certain number of them is adequate. As mentioned before, this ‘rota’ between training / testing vectors reduces the influence that outliers have on the classification score. The final classification rate is provided in the form $A \pm \beta$, where $A$ denotes the mean value of the classification scores derived from the different combinations of the testing / training vectors and $\beta$ denotes the standard deviation.

As mentioned in section 7.1, each one of the ten classes of the database consists of ten recordings, where seven of them are used as training and three of them as test vectors (i.e. 70-30 over all the ten classes). This ratio, between test and training vectors, is common amongst recognition systems. It is possible to do other combinations such as 80-20 or even 90-10, that would provide higher classification scores as the number of the training vectors is higher. However, the aim of this research is to show that phase conveys information that is useful for frequency-based feature extraction, rather than trying to achieve the highest possible classification rate. Hence, in the 70-30 random split, there are a total of 120 possible triplets of test vectors since each class consists of ten samples. However, it is not necessary to calculate all the possible combinations and hence, for these experiments 70 out of the 120 possible combinations are evaluated.

As mentioned, the aim of the aforementioned resampling method is to estimate the standard deviation of the classification rate. However, it is a computationally expensive process and thus, it is only applied to the experiments that correspond to the final classification rates presented in table 7.3.
Chapter 7. Experimental results for the audio classification application

7.2.4 Significance (Matched-pairs t-test)

Significance is a statistical measurement that indicates whether the difference between two classification scores is substantial. Specifically, it compares the performance of two classification algorithms, based on the classification rate each of them obtains [65]. Thus, this statistical measurement indicates whether the classification score obtained from the combination of the feature vectors formed from the phase related spectrograms with the feature vectors formed from the magnitude spectrograms of the signals, is significantly higher compared to the case where the feature vectors used are formed only from the magnitude spectrograms of the signals (section 7.3). In the following part, the theory behind the test (matched-pairs) selected in order to indicate whether there is significance in the difference between the classification scores obtained by two algorithms, is described.

Matched-pairs t-test

As mentioned, the aim of the significance test is, given the classification scores obtained by two algorithms $A_1$ and $A_2$ with the same set of $N$ test vectors, to decide whether the difference of their scores, is substantial. The significance test is based on the mathematical theory of hypothesis testing. Hence, the null hypothesis, $H_0$, is considered when the classification rates obtained from the two algorithms are similar (i.e. not significantly different). Consequently, in practice, the aim of the significance test is to show whether the null hypothesis $H_0$, is accepted or rejected.

Assume that algorithm $A_1$ makes $n_1$ errors (i.e. misclassifies $n_1$ test vectors) and algorithm $A_2$ makes $n_2$ errors then, the corresponding maximum likelihood error estimates are given by:

$$
\hat{e}_1 = \frac{n_1}{N} \quad (7.1) \text{ and } \\
\hat{e}_2 = \frac{n_2}{N} \quad (7.2),
$$
for \( A_i \) and \( A_j \), respectively. Moreover, the variances, \( \sigma_i^2 \) and \( \sigma_j^2 \), of \( \hat{e}_i \) and \( \hat{e}_j \), are given by:

\[
\sigma_i^2 = \frac{\hat{e}_i(1-\hat{e}_i)}{N} \quad (7.3)
\]

\[
\sigma_j^2 = \frac{\hat{e}_j(1-\hat{e}_j)}{N} \quad (7.4),
\]

respectively.

Subsequently, based on equations (7.1) and (7.2), the difference between the maximum likelihood estimates of the errors, \( \hat{e}_i \) and \( \hat{e}_j \), is given by:

\[
d = |\hat{e}_i - \hat{e}_j| \quad (7.5).
\]

Assuming that the process \( A_i \) is independent of the process \( A_j \), then, the variance of \( \hat{d} \), is given by:

\[
\sigma_d^2 = \sigma_i^2 + \sigma_j^2 \quad (7.6).
\]

Combining equations (7.3), (7.4) and (7.6),

\[
\sigma_d = \sqrt{\frac{\hat{e}_i(1-\hat{e}_i)}{N} + \frac{\hat{e}_j(1-\hat{e}_j)}{N}} \quad (7.7)
\]

Assuming that the number of the test vectors, \( N \), is large, the distribution of \( \hat{d} \) is considered to be normal. Thus, the 95.0% of the values of \( \hat{d} \) are within the range from

\[
p = \hat{d} - 2\sigma_d \quad \text{to} \quad p' = \hat{d} + 2\sigma_d.
\]

From [65], when \( p > 0 \), the null hypothesis, \( H_0 \), is rejected and hence, the difference of the two algorithms \( A_i \) and \( A_j \), in terms of their classification scores, is significant with a 95% confidence interval. Otherwise stated, if \( p > 0 \) then, with a 95.0% confidence, the difference between the maximum likelihood
estimates of the errors $\hat{\epsilon}_1$ and $\hat{\epsilon}_2$ is significant and thus, the performance, in terms of classification rate, of the algorithm $A_1$ is significantly better/worse compared to the algorithm $A_2$.

As mentioned, one of the assumptions on which this test is based, states that the number of the test vectors $N$ should be large. Note that, although the number of the test vectors used for our experiments is not large, the significance test is still an indication of the difference in the performance, in terms of classification, between the two classification methods presented in section 7.3.

7.3 Classification results of the combinatory scheme

As mentioned in the introduction of section 7.2, the classes that perform worse, in terms of classification rate, based on the feature vectors extracted from the Fourier magnitude spectrogram are: the second, the fourth, the fifth and the tenth. Thus, for these classes, where the Fourier magnitude spectrogram under-performs, the feature vectors extracted from the Hartley magnitude spectrogram and the Hartley phase 'difference' spectrogram(s) are used, as assistive feature streams (subsection 7.2.1), based on the decision rule stated in subsection 7.2.2.

Hence, in the remaining part, the results of the two experiments (subsection 7.2.2) are provided (table 7.3). The first experiment presents the classification rate obtained from the combination of the feature vectors derived from the Fourier magnitude spectrogram, the Hartley magnitude spectrogram and the Hartley phase 'difference' spectrogram via the z-transform, based on the decision rule stated in subsection 7.2.2. Then, the classification rate obtained from this combinatory scheme is compared to the classification rate obtained based on the feature vectors derived from the Fourier magnitude spectrogram only. The second experiment presents the classification score obtained from the combination of the feature vectors derived from the Fourier magnitude spectrogram, the Hartley magnitude spectrogram and the Hartley phase 'difference'
spectrogram via the DTHT. Then, as with the first experiment, the classification score obtained from this combinatory scheme is compared to the score obtained from the feature vectors extracted from the Fourier magnitude spectrogram only.

<table>
<thead>
<tr>
<th>classification score in (%)</th>
<th>1st experiment</th>
<th>2nd experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>overall (for classes 2, 4, 5 &amp; 10)</td>
<td>51.7 ± 5.4</td>
<td>95.8 ± 4.2*</td>
</tr>
<tr>
<td>overall (for all 10 classes)</td>
<td>67.2 ± 4.6</td>
<td>85.5 ± 2.7</td>
</tr>
</tbody>
</table>

Table 7.3: Classification scores in % (for classes 2, 4, 5 & 10 and for all 10 classes) derived from the combinatory scheme

(*significant, 95.0% confidence interval, matched-pairs t-test)

The classification rate obtained from the feature vectors extracted from the Fourier magnitude spectrogram, overall for the second, the fourth, the fifth and the tenth class, is 51.7% (± 5.4)% (1st column, table 7.3). Based on the decision rule (subsection 7.2.2), for the 1st experiment, combining all three feature streams (i.e. the Fourier magnitude spectrogram, the Hartley magnitude spectrogram and the Hartley phase ‘difference’ spectrogram via the z-transform) the classification rate reaches 95.8% (± 4.2%). Hence, for these four classes, there is a 44.1% improvement when compared to the classification rate obtained from the feature vectors extracted from the Fourier magnitude spectrogram only. The increase in the classification rate, after the use of the assistive feature streams, for all the tested combinations of the cross-validation (subsection 7.2.3) method, is significant (subsection 7.2.4).
For the second experiment, combining the three feature streams i.e. the Fourier magnitude spectrogram, the Hartley magnitude spectrogram and the Hartley phase 'difference' spectrogram via the DTHT, the classification rate obtained, for the four classes, reaches 93.0% (±3.1%). As with the first experiment, the classification rate obtained after the use of the additional feature streams, is significantly higher when compared to the rate obtained from the feature vectors derived from the Fourier magnitude spectrogram only. A summary of these results, for both experiments, is presented in table 7.3.

Hence, for the four classes, the classification rate obtained when the principal feature stream is assisted by the additional (assistive) feature streams (Hartley magnitude and Hartley phase), is significantly (subsection 7.2.4) higher when compared to the case where the rate is obtained based on the principal feature stream (Fourier magnitude) only, for both experiments (1st / 2nd experiment) (table 7.3). Note that, the increase in the classification rate is significant for all the tested combinations of the cross-validation (subsection 7.2.3) method.

Moreover, for all the ten classes, the increase in the classification rate obtained when the principal feature stream is assisted by the additional (assistive) feature streams (Hartley magnitude and Hartley phase), for both experiments (1st / 2nd experiment), is higher (but not significantly higher) when compared to the rate obtained from the principal feature stream (Fourier magnitude) only (2nd row, table 7.3). Specifically, for the 1st experiment, the classification rate is increased by 18.3% while for the 2nd experiment by 16.7%, for the case where the principal feature stream is combined with the assistive feature streams when compared to the rate obtained from the principal feature stream only.

### 7.4 Observations from the phase related spectrograms

There are four issues that have to be considered before extracting the statistical features from the phase spectrograms. It is important to note that this is the first time that features have been extracted from the phase spectrum for the audio classification application and
consequently, there are some points that affect the classification rate and are not mentioned in other classification related literature [93], [96].

The first issue is whether the phase related content of the signal is more efficiently encapsulated, in terms of classification performance, in the phase spectrum or in the ‘difference’ of the phase spectrum (section 4.5, chapter four). Experimental results indicate that the feature vectors extracted from the ‘difference’ of the phase spectrograms of the signals [29] provide higher classification rates, when compared to the case where the feature vectors are extracted from the phase spectrograms of the signals without calculating their ‘difference’ (subsection 7.4.1).

The second issue is related to the effect that the compensation / removal of the discontinuities from the phase spectrum has to the classification rate. Experimental results indicate that the features vectors extracted from the Fourier / Hartley spectrograms derived via the DTFT/HT approach provide higher classification scores in the case where the discontinuities are compensated in the phase spectrum. Similarly, experimental results indicate that the feature vectors extracted from the Fourier / Hartley spectrograms derived via the z-transform approach, provide higher classification scores in the case where the discontinuities are removed from the phase spectrum. However, in the cases where the Fourier / Hartley phase spectrograms are implemented via the z-transform, the number of the discontinuities that are removed from the phase spectrum depends on the width of the exclusion ‘ring’ (subsection 5.1.3, chapter five). Nevertheless, the wider the ‘ring’ width, the higher the amount of the information loss becomes. Thus, there is a trade-off between the number of the discontinuities removed from the phase spectrum and the amount of information loss. Observations related to this issue are presented in subsection 7.4.2.

The third issue is related to the effect that the sampling frequency has, on the classification rate (subsection 7.4.3).
Finally, the fourth issue is related to the comparison of the Fourier based, to the Hartley based, phase spectrograms, in terms of the classification rate that they obtain (subsection 7.4.4).

Note that many of the percentages that will be presented in the following subsections are taken from table 7.2 (section 7.2) that shows the classification rates, overall the ten classes of the database (section 7.1), derived from the signal representations.

7.4.1 The use of phase 'difference' for feature extraction from the phase spectrograms

Preliminary experimental results show that a classification rate improvement of 13.3%, overall for the ten classes is obtained, in the case where the statistical measurements of the feature vectors, are extracted from the Hartley phase ‘difference’ (i.e., the first phase point is subtracted from the second phase point, etc. for each row of the phase spectrogram along the discrete frequency axis) spectrograms (section 4.5, chapter four) compared to the case where they are extracted from the Hartley phase spectrograms (Hartley phase spectrogram via the z-transform and Hartley phase spectrogram via the DTHT) without the use of the ‘difference’. The aforementioned results indicate that the information content of the phase related spectrograms is presented to the classifier, in an improved manner, when the ‘difference’ is used. Similar observations have been made in [29] where it was observed that the phase ‘difference’ is less affected by phase noise when compared to the phase. Similarly for the Fourier case (Fourier phase spectrogram via the z-transform and Fourier phase spectrogram via the DTFT), a classification rate improvement of 16.7% overall for the ten classes is obtained, when the statistical measurements of the feature vectors are extracted from the Fourier phase ‘difference’ spectrograms (section 4.5, chapter four) when compared to the case where they are extracted from the Fourier phase spectrograms (no use of ‘difference’) [96]. Thus, the classification rates of the phase spectrograms, presented in this chapter, correspond to the case where the statistical measurements, that form the feature vectors, are extracted from the phase ‘difference’ spectrograms.
7.4.2 Variation of the classification rate with respect to the compensation / removal of the discontinuities from the phase spectrograms

In this subsection, the variation of the classification rate with respect to the compensation / removal of the discontinuities from the Hartley phase and the Fourier phase spectrograms will be presented.

For the case of the Fourier phase 'difference' spectrogram via the DTFT (based on algorithm 4 – section 5.5 – chapter five), the highest classification score is obtained when the feature vectors are extracted from the spectrograms after both the ‘extrinsic’ and the ‘intrinsic’ discontinuities are compensated when compared to the cases where only the ‘intrinsic’ or only the ‘extrinsic’ or none of the discontinuities are compensated [96]. Thus, for the case of the Fourier phase ‘difference’ spectrogram via the DTFT, the classification rate overall for the ten classes, is 50.0%, when both the ‘extrinsic’ and the ‘intrinsic’ discontinuities are compensated. Similarly, for the case of the Hartley phase ‘difference’ spectrogram via the DTHT (based on algorithm 5 – section 5.5 – chapter five), the highest classification score is obtained when the feature vectors are extracted from the spectrograms after the ‘intrinsic’ discontinuities are compensated (note that the Hartley phase spectrum does not convey ‘extrinsic’ discontinuities). Hence, for the case of the Hartley phase ‘difference’ spectrogram via the DTHT, the classification rate, overall for the ten classes, reaches 56.7% (table 7.2, section 7.2), when the ‘intrinsic’ discontinuities are compensated.

As described in subsection 5.1.3 (chapter five), the ‘sharp zeros’ (i.e. ‘zeros’ located on and very close to the circumference of the unit circle) cause discontinuities in the phase spectrum. Hence, the removal of these ‘sharp zeros’ (via the exclusion ‘ring’) reduces the discontinuities in the phase spectrum [39] and consequently, improves the classification rate, compared to the case where all the ‘zeros’ are used for the implementation of the phase spectrum. However, the wider the exclusion ‘ring’, the higher the number of ‘zeros’ excluded and consequently, the higher the information loss.
Experimental simulations showed that a 'ring' width of 0.001 excludes, on average, 15.0% of the 'zeros' of the signal (audio recordings of the database (section 7.1)) in the z-plane. Thus, the maximum 'ring' width was limited to 0.001, as a further increase in its 'width', would produce unreliable experimental results, due to the limited number of 'zeros' taken into consideration for the construction of the phase spectrograms. Note that, when the 'ring' width is equal to 0, it represents the case in which all the 'zeros' are included in the implementation of the phase spectrum. Hence, for the case of the implementation of the Hartley phase 'difference' spectrogram via the z-transform (based on algorithm 2 – section 5.5 – chapter five), the experimental results showed that when the 'ring' width is equal to zero, the classification rate, overall for the ten classes, equals 56.7%. However, the rate eventually increases, as the 'ring' width increases, obtaining its highest value, 60.0% (table 7.2, section 7.2), when the 'ring' width reaches 0.00003. Moreover, as the 'ring' width further increases from 0.00004 to 0.001, the classification score eventually decreases, obtaining its lowest rate, which is 53.3% for the widest value of the 'ring' (0.001).

Hence, the highest classification score is reached when the signal's 'zeros' which are located very close to or on, the circumference of the unit circle ('ring' width of 0.00003), are removed. Experimental simulations showed that a 'ring' width of 0.00003 excludes, on average, 1.0% of the 'zeros' in the z-plane and this gives the highest classification rate.

In theory, the highest classification score should be obtained in the case where only the 'zeros' located on the circumference of the unit circle ('intrinsic' discontinuities) are excluded (i.e. minimum possible width of the exclusion 'ring'). However, in the discrete domain, it is very rare for the 'zeros' to be located exactly on the circumference of the unit circle. Moreover, due to the limited sampling rate, the phase transitions are caused by the 'zeros' located very close to the circumference of the unit circle, as well as, by the 'zeros' located on the circumference of the unit circle [39] (subsection 5.1.3, chapter five). Consequently, the highest classification score is obtained in the case where an exclusion 'ring' with a certain width is tested.
Chapter 7. Experimental results for the audio classification application

The same performance of the classification rate, with respect to the ‘ring’ width, is observed for the case where the feature vectors are extracted from the Fourier (based on algorithm 1 – section 5.5 – chapter five), rather than the Hartley phase ‘difference’ spectrograms via the z-transform. Again, the highest classification score is reached with the same ‘ring’ width (i.e. 0.00003) [96].

7.4.3 Variation of the classification rate with respect to the sampling frequency of the phase spectrograms

A phase spectrogram (see introduction of this chapter), is implemented by dividing the time signal into frames, evaluating the phase spectrum of each frame and then by forming a matrix where each row corresponds to the phase spectrum derived from each consecutive frame. An issue that affects the classification rate is the sampling frequency, i.e., the number of the frequency points/bins of the phase spectrum. As a reminder (subsection 4.1.2, chapter four), a time signal can be fully represented in the frequency domain, where the number of frequency points is equal to the number of discrete-time points.

The classification rates obtained based on the phase spectrograms, indicate that the classification score reaches its highest value where the length of the phase spectrum is similar to the length of the time frame (i.e. 256 points). Thus, neither oversampling nor undersampling of the frequency domain, improves the classification score. Undersampling results to information suppression and consequently, reduces the classification rate whereas, oversampling just improves the resolution of the spectrum however, it ‘confuses’ the classifier by adding redundant information. This observation holds for both the Fourier and the Hartley phase spectrograms. Thus, the classification rates of the phase spectrograms presented in table 7.2 and elsewhere in this chapter correspond to the case where the length of the phase spectrum per frame, is equal to the length of the corresponding frame in the time domain (i.e. 256 points).
7.4.4 Comparison of the Hartley phase ‘difference’ with the Fourier phase ‘difference’ spectrograms in terms of classification.

In this subsection a comparison is presented, in terms of classification rate, between the Hartley phase ‘difference’ spectrograms via the DTHT and via the z-transform (based on algorithms 5 and 2 – section 5.5 – chapter five) with the Fourier phase ‘difference’ spectrograms via the DTFT and via the z-transform (based on algorithms 4 and 1 – section 5.5 – chapter five), respectively.

The experimental results show that the classification rate obtained from the Hartley phase ‘difference’ spectrogram via the DTHT is 6.7% higher, overall for the ten classes, compared to the rate obtained from the Fourier phase ‘difference’ spectrogram via the DTFT. As a reminder, for the Fourier case, the construction of the phase spectrum can be summarized in three stages. The initial stage is the evaluation of the phase spectrum based on equation (5.1) of chapter five, the next stage is the application of the ‘unwrapping’ algorithm (removal of the ‘extrinsic’ discontinuities) to the phase spectrum, while the last stage is the application of the \( \pm \pi \) compensation algorithm (removal of the ‘intrinsic’ discontinuities). However, the construction of the Hartley phase spectrum can be summarized in two stages. The initial stage is the evaluation of the Hartley phase spectrum based on equation (5.12) (section 5.3, chapter five) and the next stage is the application of the compensation (multiplication with \(-1\)) algorithm (removal of the ‘intrinsic’ discontinuities). The stage related to the compensation of the ‘extrinsic’ discontinuities is omitted because the Hartley phase spectrum does not convey these discontinuities (section 5.3, chapter five). Hence, this 6.7% difference in the classification rate, is explained from the fact that the ‘unwrapping’ algorithm, applied to the Fourier phase spectrum in order to remove the ‘extrinsic’ discontinuities, causes ‘wrapping’ ambiguities that affect its performance, in terms of classification.

More importantly, experimental results show that the classification rate obtained from the Hartley phase ‘difference’ spectrogram via the z-transform is, on average for each ‘ring’ width (subsection 7.4.2), 3.3% higher overall for the ten classes, when compared to the
classification rate obtained from the Fourier phase 'difference' spectrogram via the z-transform. This is not a significant increase but it may be an indication of improved performance. Hence, this difference in the classification rate indicates that even in the case where the 'extrinsic' discontinuities do not exist (neither the Hartley phase nor the Fourier phase spectra derived via the z-transform, have 'extrinsic' discontinuities, see subsection 5.1.3 and section 5.3 of chapter five), the Hartley phase 'difference' spectrogram still outperforms the Fourier phase 'difference' spectrogram. Thus, the Hartley phase 'difference' spectrogram presents the phase related content of the signal to the classifier in an improved manner, when compared to the Fourier phase 'difference' spectrogram. Consequently, this indicates that the Hartley phase spectrum (section 5.3, chapter five), encapsulates the signal's phase content in an improved manner when compared to the Fourier approach (section 5.1, chapter five). The classification rates of the phase spectrograms, presented in table 7.2 (over all the ten classes), correspond to the case where the statistical measurements, that form the feature vectors, are extracted from the Hartley phase 'difference' spectrograms (via the D'THT and via the z-transform) rather than from their Fourier counterparts.

Summarizing, the four observations, presented in this section, indicate that the phase spectrograms perform better, in terms of classification rate, in the case where the statistical features that form the feature vectors are extracted using the 'difference' (subsection 7.4.1), when the discontinuities are compensated / removed (subsection 7.4.2) and when the length of the each frame (row) of the phase spectrogram is the same compared to the length of the time frame in which the signal is divided (subsection 7.4.3). Moreover, the fourth observation, presented in subsection 7.4.4, indicates that the phase spectra based on the Hartley approach outperform the phase spectra based on the Fourier approach.
7.5 Summary and conclusions

Summarizing, the experimental results of this chapter show that the combination of the magnitude with the phase-related feature vectors (section 7.3) provide improved performance, in terms of classification rate, when compared to the case where only the magnitude feature vectors are used. Hence, for certain classes, in which the Fourier magnitude spectrogram under-performs, the use of the phase-related spectrograms increases the classification rate, significantly. Consequently, the features extracted from the phase-related spectrum can form feature vector(s) that are effective for the application of frequency based statistical feature extraction for audio classification. In section 7.4, useful observations related to the efficient feature extraction from the phase spectrograms, are stated.

The proposed scheme is tested to a database of audio recordings of gunshots (section 7.1). The classes of gunshots are acoustically similar ('fine' classification) and hence, the classification task is more demanding compared to the case where the classes of the database were acoustically dissimilar (e.g. discrimination between sports sounds – 'coarse' classification).

The audio classification application (chapter seven) together with the time localization of a signal application (chapter six) indicate that the phase content of a signal is encapsulated more efficiently in the Hartley phase rather than in the Fourier phase function. More importantly, the last two chapters show that the phase function encapsulates information that is distinct from the magnitude information and also, its usefulness is shown when applied to tasks involving either synthetic (chapter six) or non-synthetic (chapter seven) signals.
CHAPTER 8
CONCLUSIONS AND FUTURE PLANS

8.1 Conclusions of the research

A major part of this thesis is focused on the study of the phase content of synthetic and non-synthetic signals. The difficulties (discontinuities) in processing the phase spectrum of a signal have probably led previous researchers to avoid its investigation. Hence, as a reminder, the first category of the discontinuities, that appear in the Fourier phase spectrum, is related to the computation of the inverse tangent function ('extrinsic' discontinuities) while, the second category, is related to properties of the signal ('intrinsic' discontinuities). The major disadvantage of the implementation of the phase spectrum based on the Fourier transform, is the heuristic approach of the compensation of the 'extrinsic' discontinuities ('wrapping' ambiguities). A novel approach to phase analysis, is the implementation of the phase spectrum based on the Hartley transform. Specifically, in the case where the phase spectrum is implemented based on the Hartley transform, it does not suffer from the 'extrinsic' discontinuities and consequently, the 'wrapping' ambiguities do not arise.
In chapter six the Hartley phase spectrum is compared to the Fourier phase spectrum, via their corresponding cepstrums, for the application of time delay estimation. The experiments, involved synthetic signals tested in a noisy environment. The simulations showed that the Hartley phase spectrum is more tolerant to noise and moreover, it can identify more than a single pulse, even in the case where noise is present. The noise immunity of the Hartley phase function, when compared to its Fourier counterpart, is explained via the shape of the probability density function of the Hartley phase spectrum, in the case where the signal in the time domain is Gaussian noise.

Another outcome from the work reported here, is a method for the successful discrimination of audio classes that are similar acoustically. Hence, the experimental results of the audio classification application (chapter seven), showed that the combination of the magnitude with the phase-related feature vectors provide improved performance, in terms of classification rate, when compared to the case where only the magnitude feature vectors are used. Specifically, for certain classes, in which the Fourier magnitude spectrogram under-performs, the assistance of the Hartley phase spectrogram increases the classification rate, significantly. Consequently, the features extracted from the phase-related spectra can form feature vectors that are effective for the application of frequency based statistical feature extraction for audio classification.

The time localization application (for synthetic signals) together with the audio classification application (for non-synthetic signals) indicate that the Hartley phase function can be used as a substitute to the Fourier phase function, as it encapsulates the phase content of the signal in a more efficient and easily identifiable manner. More importantly, both applications show that the phase spectrum encapsulates information that is distinct from the magnitude information and also, its usefulness is shown when applied to tasks involving either synthetic (chapter six) or non-synthetic (chapter seven) signals.
8.2 Future plans of the research

In this section, three possible future directions of the research are presented. The first suggestion is related to the ‘root’ finding algorithm employed for the evaluation of the phase spectrum via the z-transform (subsection 5.1.3, chapter five and section 4.3, chapter four). The second recommendation is related to the dimensionality reduction of the information conveyed to the phase related spectrograms (Introduction of chapter seven and section 4.4, chapter four) for the application of audio classification, while the third recommendation is related to the application of the Hartley phase spectrum to other research areas of signal processing.

8.2.1 ‘Root’ finding method of high-order polynomials

As mentioned, in chapter five, the phase evaluation via the z-transform (subsections 5.1.3 and 5.3.2 of chapter five) is based on the estimation of the ‘roots’ of high-order polynomials. However, the ‘roots’ finding methods do not always evaluate the ‘roots’ of a polynomial, accurately. A method that should provide more accurate values of the ‘roots’ of the polynomial(s) is the Lindsey-Fox Grid-Search algorithm [77]. According to this method, the polynomial is evaluated over a large grid in the z-plane (complex spectrum). A ‘root’ exists whenever a location of the grid corresponds to a local minimum value of the polynomial. More importantly, the grid can be become ‘finer’ (i.e. more dense) around the circumference of the unit circle where the majority of the ‘roots’ of the polynomial(s) formed from non-synthetic signals (figure 5.4 of chapter five), are located.

8.2.2 Dimensionality reduction of the phase related spectrograms

As mentioned in section 4.4 of chapter four, the phase related spectrograms are characterized by many discontinuities and hence, it is not possible to reduce their dimensionality via the application of another transform. However, the extraction of statistical measurements employed for the audio classification application (chapter seven)
is less efficient when compared to a dimensionality reduction method, which is based on the selection of coefficient(s) evaluated via another transform. Consequently, a possible future direction is the application of a smoothing technique (e.g. splines) on the phase spectrogram before applying a transform with compression qualities (e.g. 2D cepstrum [59] and [94]).

8.2.3 Applications of the Hartley phase spectrum

In this thesis, the attractive properties of the Hartley phase spectrum, when compared to its Fourier counterpart, are demonstrated via the applications of time localization of a signal (chapters five and six) and audio classification (chapter seven). Moreover, the properties of the Hartley phase spectrum are also useful for the speech coding application as shown in [63] and [64].

Other possible scientific areas for the application of the Hartley phase spectrum are the frequency-based system identification and speech recognition via the 'root' cepstrum [60]. The 'root' cepstrum provides an interesting alternative to the complex log cepstrum, whereby the log and antilog (inverse process) operators of the complex log cepstrum are replaced by \((\gamma)^\gamma\) and \((\gamma)^{-\gamma}\), where \(\gamma\) is constrained to lie between \(-1 \leq \gamma \leq 1\). The advantage of the 'root' cepstrum, compared to the complex log cepstrum is its ability to model a wider range of systems. However, in order to construct a cepstrum, the phase has to be 'unwrapped' thus requiring the use of a phase unwrapping algorithm ('wrapping' ambiguities). The Hartley phase spectrum could be employed so as to evaluate the phase spectrum of the signal without suffering from 'wrapping' related ambiguities.
APPENDICES

Appendix A1

The eigenvalue 'roots' finding method is based on a theorem, which states that the 'roots' of the characteristic polynomial \( p(\lambda) \) are the eigenvalues of its companion matrix \( C \).

The characteristic polynomial, \( p(\lambda) \) of \( C \), is equal to: \( p(\lambda) = \Delta(C - \lambda I_n) \), where \( \Delta \) denotes determinant and \( I_n \) denotes the \( n \times n \) size unit matrix. Hence, the 'roots' of a polynomial can be found by following these steps:

a) form a matrix \( C \) (companion matrix) which has as characteristic polynomial the function \( p(\lambda) \).

b) evaluate the eigenvalues of \( C \).

Assuming that \( p(t) = c_0 + c_1 t + \ldots + c_n t^{n-1} + t^n \) (equation (4.14), chapter four) is the characteristic polynomial, then the companion matrix, \( C \), is formed as:

177
Note that the problem of the calculation of the ‘roots’ of the polynomial \( p(t) \) (equation (4.14), chapter four), has now been reduced to the estimation of the eigenvalues of the matrix \( C \). The method chosen, in order to evaluate the eigenvalues, is the \( QR \) method [76] and is based on the \( QR \) decomposition of \( C \). The name \( QR \) is formed from the use of the letters \( Q \) and \( R \) that stand for orthogonal and right triangular, respectively. According to matrix theory, any matrix can be factorized into the product of a matrix with orthonormal columns \( (Q) \) multiplied with a matrix that is non-zero only in its upper or right triangle \( (R) \) [76].

The \( QR \) method can be summarized in the following steps:

Step 1: Assuming, \( A = Q_0 R_0 \) is a \( QR \) factorization of the matrix \( A \), then evaluate \( A_i = R_0 Q_0 \).

Step 2: Assuming, \( A_i = Q_i R_i \) is a \( QR \) factorization of the matrix \( A_i \), then evaluate \( A_2 = R_1 Q_1 \).

Step 3: Repeat the same procedure i.e. assuming, \( A_m = Q_m R_m \) is a \( QR \) factorization of the matrix \( A_m \), then evaluate \( A_{m+1} = R_m Q_m \).

The process terminates when the values below the main diagonal of the matrix, \( A_m \), are very small or if convergence does not occur.
Each iteration results in a matrix, $A_{m+1} = R_m Q_m$, that eventually approaches an upper triangular matrix while its eigenvalues ('roots' of $p(t)$) appear in the main diagonal.
Appendix A2

Distribution formed by the Gaussian noise in the Hartley phase spectrum

As mentioned in section 5.3 (equation (5.12)) of chapter five, the Hartley phase function is defined as the ratio of the Hartley spectrum over the Fourier magnitude i.e.,

\[ u = \frac{x + y}{\sqrt{x^2 + y^2}} \quad (6.1A) \]

where, \( x \) and \( y \) represent the real and the imaginary components of the Fourier transform of the signal, respectively. The objective of this appendix is to specify the probability density function (pdf) of \( u \), in the case where the signal in the time domain is Gaussian distributed noise.

Note that, the Hartley transform can be defined either as \( H(\omega) = \Re(S(\omega)) - \Im(S(\omega)) \) or \( H(\omega) = \Re(S(\omega)) + \Im(S(\omega)) \) and hence, the numerator of equation (6.1A) could be either \( x - y \) or \( x + y \), respectively. For simplicity, the latter definition is chosen.

Proof

The real, \( x \), and the imaginary, \( y \), components of the Fourier transform of Gaussian noise, are Gaussian distributed as well. Hence, the numerator of equation (6.1A) forms a Gaussian distribution (\( x \) and \( y \) are Gaussian distributed) whereas, the denominator forms a Rayleigh distribution [50]. Consequently, the aim is to derive the pdf of the division of a Gaussian distribution by a Rayleigh distribution.

Assume a coordinate system that is \( \frac{\pi}{4} \) radians tilted with respect to the conventional Cartesian coordinate system. \( \Lambda(x,y) \) is a randomly chosen pair located on the original...
coordinate system and point B is the projection of point A on the new (tilted) coordinate system, as shown in figure 6.1A.

![Figure 6.1A: Coordinate system, \( \frac{\pi}{4} \) radians tilted with respect to the conventional Cartesian system]

\( A(x, y) \)

\( \theta \)

\( B \)

\( C \)

\( M \)

\( N \)

\( V \)

\( n \)

\[ A(x, y) \]

\[ \theta \]

\[ B \]

\[ C \]

\[ M \]

\[ N \]

\[ V \]

\[ n \]

\[ x^2 \]

\[ y^2 \]

\[ \sqrt{x^2 + y^2} \]

\[ \sqrt{2} \]

\[ \cos \phi \]

\[ BC \]

\[ AC \]

\[ \cos \phi = \frac{BC}{AC} = \frac{x + y}{\sqrt{x^2 + y^2}} = \frac{x + y}{\sqrt{2}(\sqrt{x^2 + y^2})} \quad (6.2A) \]

From equations (6.1A) and (6.2A),

\[ u = \sqrt{2} \cos \phi \]

Consequently, the distribution of \( \sqrt{2} \cos \phi \), rather than the distribution of \( \frac{x + y}{\sqrt{x^2 + y^2}} \), has to be specified.

The method followed, to obtain the distribution of \( \sqrt{2} \cos \phi \), is summarized as follows:

\[ -1 < \cos \phi < 1 \leftrightarrow \]

\[ -\sqrt{2} < \sqrt{2} \cos \phi < \sqrt{2} \leftrightarrow \]
Assume that, $\beta = \sqrt{2} \cdot \cos \phi$.

Hence, the probability of $\beta$ to obtain a certain value $\beta_o$, is given by:

$$p(\beta_o) = \lim_{\Delta \to 0} \frac{P(\beta_o < \beta < \beta_o + \Delta)}{\Delta}, \text{ but because } \phi \text{ is uniformly distributed between 0 and } 2\pi \text{ then,}$$

$$p(\beta_o) = \lim_{\Delta \to 0} \frac{1}{2\pi} \frac{P(\beta_o < \beta < \beta_o + \Delta)}{\Delta} \quad (6.3A)$$

The value of $P(\beta_o < \beta < \beta_o + \Delta)$ equals to the number of the angles that correspond to the range from $\beta_o$ to $\beta_o + \Delta$.

Thus,

$$\beta_o < \beta < \beta_o + \Delta \iff \beta_o < \sqrt{2} \cos \phi < \beta_o + \Delta \iff$$

$$\frac{\beta_o}{\sqrt{2}} < \cos \phi < \frac{\beta_o + \Delta}{\sqrt{2}} \iff$$

$$ar \cos \frac{\beta_o}{\sqrt{2}} > \phi > ar \cos \frac{\beta_o + \Delta}{\sqrt{2}} \text{ for each interval, from 0 to } \pi \text{ and from } \pi \text{ to } 2\pi.$$

Note that $\beta$ is located on the horizontal axis of the trigonometric circle. Consequently, the number of angles that correspond to the interval between $\frac{\beta_o}{\sqrt{2}}$ and $\frac{\beta_o + \Delta}{\sqrt{2}}$ have to be multiplied by two, as both intervals from, 0 to $\pi$ and from $\pi$ to $2\pi$, have to be considered.

Thus, the numerator, $P(\beta_o < \beta < \beta_o + \Delta)$, of equation (6.3A) equals to:
2. \( (\arccos \frac{\beta + \Delta}{\sqrt{2}} - \arccos \frac{\beta}{\sqrt{2}}) \).

Substituting to equation (6.3A),

\[
p(\beta_n) = \lim_{\Delta \to 0} \frac{1}{2\pi} \cdot \frac{2. (\arccos \frac{\beta + \Delta}{\sqrt{2}} - \arccos \frac{\beta}{\sqrt{2}})}{\Delta}
\]

(6.4A)

Moreover, for any function \( p \), the following relationship holds,

\[
p(\beta + \Delta) - p(\beta) = p'(\beta) \Delta,
\]

(6.5A)

where \( p' \) is the first derivative of the function \( p \).

Hence, from equation (6.4A),

\[
p(\beta) = \frac{1}{\pi} \arccos \left( \frac{\beta}{\sqrt{2}} \right)
\]

Consequently,

\[
p(\beta) = \frac{1}{\pi \sqrt{2}} \sqrt{1 - \left( \frac{\beta}{\sqrt{2}} \right)^2}
\]

(6.6A)

which is the probability density function required (equation (6.2), chapter six).
BIBLIOGRAPHY

[1] Zhang, T. and Kuo, C.C.J.
Audio content analysis for online audiovisual data segmentation and classification.

Content-based retrieval of music and audio.

Audio feature extraction and analysis for scene classification.
Pages:343 – 348.

Hierarchical classification of audio data for archiving and retrieving.
Pages:3001 - 3004 vol.6.

Audio-guided audiovisual data segmentation, indexing and retrieval.
[6] Li, S.Z.  
Content-based audio classification and retrieval using the nearest feature line method.  

Multimedia content analysis using both audio and visual clues.  
Pages: 12 – 36.

Audio Classification.  
*ASSAVID project*, University of Surrey, May 2001.

Audio Signal Classification: An Overview.  
Canadian Artificial Intelligence. Winter 2000. 4-6.

A study on content-based classification and retrieval of audio database.  

Video scene determination using audiovisual data analysis.  
Pages: 124 – 129.

Pitch-based feature extraction for audio classification.  
Pages: 43 – 47.

Classification of audio signals using statistical features on time and wavelet transform domains.  
Pages: 3621 - 3624 vol.6.

Content based audio classification and retrieval using joint time-frequency analysis.  
Pages: V - 665-8 vol.5.
Joint audio visual retrieval for tennis broadcasts.

[16] Casey, M.
MPEG-7 sound-recognition tools.

[17] Hyoung-Gook Kim; Moreau, N. and Sikora, T.
Audio classification based on MPEG-7 spectral basis representations.

[18] Xiong, Z.; Regunathan Radhakrishnan and Ajay Divakaran.
Generation of sports highlights using motion activity in combination with a common audio feature extraction framework.

[19] Regunathan Radhakrishnan; Xiong, Z.; Ajay Divakaran, and Ishikawa, Y.
Generation of sports highlights using a combination of supervised & unsupervised learning in audio domain.

[20] Leonardi, R.; Migliorati, P. and Prandini, M.
Semantic indexing of sports program sequences by audio-visual analysis.

[21] Kijak, E.; Gravier, G.; Gros, P.; Oisel, L. and Bimbot, F.
HMM based structuring of tennis videos using visual and audio cues.

[22] Ziyou Xiong; Radhakrishnan, R.; Divakaran, A. and Huang, T.S.
Audio events detection based highlights extraction from baseball, golf and soccer games in a unified framework.
[23] Wold E.; Blum T., etal.  
Content-Based classification, search and retrieval of audio. 

Audio segmentation and classification based on a selective analysis scheme. 

Continuous, Speaker Independent, Speech Recognition for a Speech to Viseme Translator.  

[26] Long, D.G.  
Exact computation of the unwrapped phase of finite-length time series.  
Pages:1787 – 1790.

[27] McGowan, R. and Kuc, R.  
A direct relation between a signal time series and its unwrapped phase.  
Pages:719 – 726.

[28] Bhanu, B. and McClellan, J.  
On the computation of the complex cepstrum.  
Pages:583 – 585.

[29] Gough, P.  
A particular example of phase unwrapping using noisy experimental data.  
Pages:742 – 744.

Calculating the complex cepstrum without phase unwrapping or integration.  
Pages:1014 – 1017.

Algebraic multidimensional phase unwrapping and zero distribution of complex polynomials-characterization of multivariate stable polynomials.  
Pages:1639 – 1664.
[32] Bonzanigo, F.
An improvement of Tribolet's phase unwrapping algorithm.  
Pages:104 – 105.

[33] Scott, J.
Improving confidence in the phase unwrapping algorithm.  
Pages:1254 – 1255.

[34] Nico, G. and Fortuny, J.
Using the matrix pencil method to solve phase unwrapping.  
Pages:886 – 888.

[35] Steiglitz, K. and Dickinson, B.
Phase unwrapping by factorization.  
Pages:984 – 991.

[36] Hayes, M.; Jae Lim and Oppenheim, A.
Signal reconstruction from phase or magnitude.  
Pages:672 – 680.

[37] Quatieri, T. Jr. and Oppenheim, A.
Iterative techniques for minimum phase signal reconstruction from phase or magnitude.  
Pages:1187 – 1193.

[38] Petropulu, A.P. and Pozidis, H.
Phase reconstruction from bispectrum slices.  
Pages:527 – 530.

[39] Al-Nashi, H.
Phase unwrapping of digital signals.  
Pages:1693 – 1702.
[40] Volker, B. and Handel, P.
Frequency estimation from proper sets of correlations.
Pages:791 – 802.

[41] Tribolet, J.
A new phase unwrapping algorithm.
Pages:170 – 177.

[42] Paliwal, K.K. and Alsteris, L.D.
On the usefulness of phase spectrum in human listening tests.
*Speech Communication*, vol. 45, Feb. 2005

[43] Bozkurt, B. and Dutoit, T.
Mixed-Phase Speech Modeling and Formant Estimation, Using Differential Phase Spectrums.

Importance of window shape for phase-only reconstruction of speech.
Pages:I - 573-6 vol.1.

[45] Ghiglia, D.C. and Pritt, M.D.
Two-Dimensional Phase Unwrapping: Theory, Algorithms and Software.

[46] Rabiner, L.
On the use of autocorrelation analysis for pitch detection.
Pages:24 – 33.

A comparative performance study of several pitch detection algorithms.
Pages:399 – 418.

[48] Makhoul, J.
Linear Prediction: A tutorial review.


[51] Li, D, Sethi; I.K, Dimitrova, N. and McGee, T.
Classification of general audio data for content-based retrieval.

[52] Manjunath, B.S. (Editor); Salembier, P. (Editor); Sikora, T. (Editor)
and Salembier, P.
Introduction to MPEG 7: Multimedia Content Description Language.

[53] Parsons, S. and Jones, G.
Acoustic identification of 12 species of echolocating bat by discriminant function analysis and artificial neural networks.
Pages: 2641-2656.

[54] Pandit, M.; Yusoff, Y.; Kittler, J.; Christmas, W.J. and Chilton, EHS.
Video assisted segmentation of speech and audio track.
*Proceedings of the Society of Photo-optical Instrumentation Engineers (SPIE)*, 3846,

[55] Pandit, M.; Kittler, J. and Christmas, W.J.
Audio Classification for annotation of sports videos.
*Proceedings of Infotech Oulu International Workshop on Information Retrieval*,

[56] Furui, S.
Cepstral analysis technique for automatic speaker verification.
Pages:254 - 272.

[57] Itakura, F.
Line spectrum representation of linear prediction coefficients of speech signal.

[58] Noll, A.M.
Cepstrum Pitch Determination.
Pages:293-309.

[59] Marvi, H.
Efficient feature extraction based on two-dimensional cepstrum analysis for speech recognition.

[60] Sarpal, S.
Efficient system identification based on root cepstral deconvolution.
[61] Atmore, K.S.
Phase analysis of speech and audio signals.

[62] Harris, F.J.
On the use of windows for harmonic analysis with the DFT.
Pages:55 – 83.

[63] Chilton, EHS.
An 8kb/s speech coder based on the Hartley transform.
ICCS '90 Communication Systems: Towards Global Integration, Vol. 1,
pp 13.5.1-13.5.5.

[64] Chilton, E.
The Hartley transform applied to speech coding.
Speech Coding, IEE Colloquium on

[65] Gillick, L. and Cox, S.J.
Some statistical issues in the comparison of speech recognition algorithms.
Conference on

[66] Proakis, J.G. and Manolakis,D.G.

[67] Millane, R.P.
Analytic properties of the Hartley transform and their implications.
Proceedings of the IEEE, Volume: 82, Issue: 3, March 1994
Pages:413 – 428.

[68] Pattern Classification.
Duda, R.O.; Hart, P.E. and Stork, D.G.

[69] Hlawatsch, F. and Boudreaux-Bartels, G.F.
Linear and quadratic time-frequency signal representations.
Signal Processing Magazine, IEEE, Volume: 9, Issue: 2, April 1992
Pages:21 – 67.

[70] Cohen, L.
Time-frequency distributions-a review.
Pages:941 – 981.

[71] Olejniczak, K.J. and Heydt, G.T.
Scanning the Special Section on the Hartley.
[72] Bracewell, R.N.
The discrete Hartley Transform.

[73] Bracewell, R.N.
The Fourier Transform and Its Applications.

[74] Ifeachor, E.C. and Jervis, B.W.
DSP: A Practical Approach.
Addison-Wesley, 1993.

[75] Bracewell, R.N.
Aspects of the Hartley Transform.


[77] Sitton, G.A.; Burrus, C.S.; Fox, J.W. and Treitel, S.
Factoring very-high-degree polynomials.
Pages: 27 – 42.

[78] Mahalanobis, P.C.
On the generalized distance in statistics.
*Proceedings of the National Institute of Science of India*, 12 (1936) 49-55.

[79] Webb, A.R.
Statistical Pattern Recognition.

[80] Mansfield, E.
Basic Statistics with Applications.

[81] Papoulis, A.
Probability and Statistics.

Statistical pattern recognition: a review.
Pages: 4 – 37.


Computer Methods for Mathematical Computations.

[86] DeFatta, D.J.; Lucas, J.G. and Hodgkiss, W.S.

[87] Chen, C-T.

[88] Nikias, C.L. and Petropulu, A.

[89] Kittler, J.
Pattern Recognition Lecture Notes.
University of Surrey, 2001.

[90] Chilton, EHS.
A Continuing Education Course on Advanced Digital Signal Processing Lecture Notes.
University of Surrey, 2001.

[91] Poularikas, A.D.
The Transforms and Applications Handbook.

[92] Oppenheim, A.V.; Schafer, R.W. and Buck, J.R.
Discrete-Time Signal Processing.

[93] Paraskevas, I. and Chilton, E.
Phase as an assistive feature vector for audio classification.
Spring Conference, Southampton, UK March 2004
Proceedings of the Institute of Acoustics, vol. 26, pt. 1,
pages 590 – 600.

[94] Marvi, H.; Paraskevas, I. and Chilton, E.
Acoustic classification using time – frequency distributions.
Spring Conference, Southampton, UK March 2004
Proceedings of the Institute of Acoustics, vol. 26, pt. 1,
pages 612 – 620.

[95] Paraskevas, I. and Chilton, E.
Audio classification using acoustic images for retrieval from multimedia databases.
Video/Image Processing and Multimedia Communications, 2003, 4th EURASIP
Conference focused on
Volume 1, 2-5 July 2003 Page(s):187 - 192 vol.1.
[96] Paraskevas, I. and Chilton, E.  
Combination of Magnitude and Phase Statistical Features for Audio Classification.  

[97] Chilton, E. and Paraskevas, I.  
Audio fine classification using the statistical analysis of acoustic images.  
*Acoustical Society of America (Nashville meeting)* - April 2003 - Volume 113, issue 4, page 2271.

[98] Chilton, E. and Paraskevas, I.  
Fine classification for acoustic images.  

[99] Hartley, R.V.L.  
A more symmetrical Fourier analysis applied to transmission problems.  
*Proc. IRE*, vol. 30, 1942, pp. 144-150.

[100] Rabiner, L.R. and Gold, B.  
Theory and Applications of Digital Signal Processing,  

Elementary Statistics.  
Prentice-Hall, 1983.

[102] Minium, E.W. and Clarke, R.B.  
Elements on statistical reasoning.  

[103] Elston, R.C. and Johnson, W.D.  
Essentials of Biostatistics.  

[104] Paliwal, K.K. and Atal, B.S.  
Frequency-related representation of speech.  

[105] Micallef, P.  
A text to speech synthesis system for Maltese.  

[106] Rabiner, L. and Juang, B.H.  
Fundamentals of speech recognition.  

[107] Rabiner, L. and Schafer, R.W.  
Digital processing of speech signals.  
The quefreny analysis of time series for echoes: Cepstrum, pseudo-auto-covariance, cross-cepstrum and saphe cracking.
*Proceedings of the Symposium on Time Series Analysis, 1963*

[109] Childers, D.G.; Skinner, D. P. and Kemeralt, R.
The cepstrum: a guide to processing.
*Proceedings of IEEE, Oct. 1967*
Pages:1428 – 1443.

[110] Davis, S.; and Mermeistein, S.
Comparison of parametric representation for monosyllabic word recognition in continuously spoken sentences.
*IEEE Transactions on Acoustic, speech and signal processing, 1990*
28:357 - 366.


[112] [http://mathworld.wolfram.com/RampFunction.html](http://mathworld.wolfram.com/RampFunction.html)

[113] Kostek, B.
Musical instrument classification and duet analysis employing music information retrieval techniques.
*Proceedings of the IEEE, Volume: 92, Issue: 4, Apr, 2004*
Page(s):712 - 729.

[114] Tzanetakis, G. and Cook, P.
Musical genre classification of audio signals.
*Speech and Audio Processing, IEEE Transactions on, Volume: 10, Issue: 5, July 2002*
Page(s):293 – 302.

[115] Moreno, I.; Kober, V.; Lashin, V.; Campos, J.; Yaroslavsky, L.P. and Yzuel, M.J.
Color pattern recognition with circular component whitening.
*Optics Letters, Volume: 21, Issue: 7, April 1996*
Page(s):498 – 500.

[116] Brown, G.J. and Cooke, M.P.
Computational auditory scene analysis.
*Computer Speech and Language, 8, 1994*
Pages(s): 297 - 336

[117] Bregman, A.S.
Auditory Scene Analysis.