Bispectral Analysis of Speech Signals

Justin W A Fackrell

A thesis submitted for the degree of Doctor of Philosophy.

The University of Edinburgh.

-September 1996-

Abstract

Techniques which utilise a signal's Higher Order Statistics (HOS) can reveal information about non-Gaussian signals and nonlinearities which cannot be obtained using conventional (second-order) techniques. This information may be useful in speech processing because it may provide clues about how to construct new models of speech production which are better than existing models.

There has been a recent surge of interest in the application of HOS techniques to speech processing, but this has been handicapped by a lack of understanding of *what* the HOS properties of speech signals are. Without this understanding the HOS information which is in speech signals can not be efficiently utilised.

This thesis describes an investigation into the use of HOS techniques, in particular the third-order frequency domain measure called the bispectrum, to speech signals. Several issues relating to bispectral speech analysis are addressed, including nonlinearity detection, pitch-synchronous analysis, estimation criteria and stationarity. A flaw is identified in an existing algorithm for detecting quadratic nonlinearities, and a new detector is proposed which has better statistical properties. In addition, a new algorithm is developed for estimating the normalised bispectrum of signals contaminated by transient noise.

Finally the tools developed in the study are applied to a specially constructed database of continuant speech sounds. The results are consistent with the hypothesis that speech signals do not exhibit quadratic nonlinearity.

Declaration of Originality

I hereby declare that the research recorded in this thesis, and the thesis itself, is the original and sole work performed by the author while studying in the Department of Electrical Engineering at The University of Edinburgh.

Justin W A Fackrell

Acknowledgements

Doing my PhD at Edinburgh University has been like dining at a fine restaurant. Thanks must go to many people for making this such an enjoyable meal, but a special *Diolch yn fawr* to my parents, who dropped me off at the door. The Maitre d', Steve McLaughlin, welcomed me as I arrived, showed me to my table and provided numerous helpful menu suggestions throughout the evening.

The appetizers, which were delicious, were provided by Paul White, Joe Hammond (at ISVR, Southampton University) and Alan Parsons (at DRA Portland), together with seasoning from Steve Tee, ISVR (*Badger*), and the Bristol90 crowd.

At adjacent tables were my fellow diners Mike Banbrook, and Achilleas Stogioglou, who helped me choose several dishes, and warned me from trying others. Other menu choices were made following advice from Steve Isard (Dept of Linguistics), Alan Wrench (CSTR), Bill Collis (ISVR), Marty Johnson (ISVR) and Peter Hill (ISVR). The occasional helpful words from Drew Lowry and Julian Page (BT Labs), Christine Shadle (Southampton University) and Paul White (ISVR) helped decipher some of the more ethereal menu items.

All the fresh meat was provided by Kim, Bill, Iain, Gillian, Dora, Paul, Emma, Jeff, Simon, Gary, Candy, Matthew, Paul, Rhona and Caroline (they provided the speech data). The SaS Group provided the vegetables, and side dishes were provided by John Dalle Molle (University of Texas), Jes Thyssen (Tele Danmark Research), Guotong Zhou (University of Virginia) and Mark Williams (Imperial College). Many of the above helped me choose (and drink) the wine. I would also like to acknowledge the assistance given to me by Orestis Papsouliotis (Statlab, University of Edinburgh) who provided advice at a critical time. After coffee, queries with the bill were settled following help from several group members - Bernie Mulgrew, Jon Altuna, Stevie Bates and Steve McLaughlin (again !).

I shared my table with Kim Hardie, who encouraged me to keep eating even when I thought I was full. She finished her meal quite some time before me, but was very patient waiting for me to finish. *Bedankt !*

Finally, this whole culinary experience would not have been possible without the help of EPSRC and BT Laboratories, Martlesham Heath, who picked up the bill, and called me a taxi.

Contents

List of Figures

Nomenclature

$B(k,l)$	discrete bispectrum
$b^2(k,l)$	discrete squared bicoherence
$\overline{b^2_{\mathrm{IT}}}$	squared bicoherence averaged over the IT
$\overline{b^2_i}(i=1,..,4)$	squared bicoherence averaged over the IT
$\sum_{\mathrm{IT}} b^2$	squared bicoherence summed over the IT
f_0	fundamental speech frequency (in Hz)
f_1, f_2	(Chapters 6 and 7) first and second formant frequencies (in Hz)
f_1, f_2, f_3	(other Chapters) normalised frequency components
f_s	sampling frequency (in Hz)
$i(n)$	discrete impulse train
k,l	discrete frequency (in $bins$)
K	number of data segments
$l(n)$	discrete Laryngograph signal
L_{IT}, L_i	number of bifrequencies in the IT or in the ith subzone of the IT
M	DFT size
m_i	ith order spectral moment (see Appendix C.3)
n	discrete time (in $samples$)
N	vector length
$N(\mu,\sigma^2)$	normal distribution with mean μ and variance σ^2
$P(k)$	discrete power spectrum
P_{FA}	probability of false alarm
$\overline{q_i}(i=1,..,4)$	average number of QPC hits in the ith subzone of the IT
t	continuous time (in s)
$s(n)$	discrete speech signal
$s^2(k,l)$	discrete skewness function
$v(n)$	discrete noise signal
$x(n)$	discrete signal
$e(n)$	iid signal

α	trimmed-means parameter
$\alpha(1)$	significance level for squared bicoherence magnitude
$\alpha(2)$	significance level for squared bicoherence phase
β	parameter for measuring peak-discrimination performance of b^2 (Section 3.7.1)

Δ	time delay (in s)
T	time (in s)
$\gamma^2(k)$	discrete second-order coherency function
μ	mean value, first-order moment about the mean [3, p49]
μ_3	skewness, third-order moment about the mean
μ_4	kurtosis, fourth-order moment about the mean
σ^2	variance, second-order moment about the mean
τ	discrete time delay (in *samples*)
$\theta, \phi, \angle[\,]$	angle (in *rad*)
χ^2_n	central χ^2 with n degrees of freedom
$\stackrel{\triangle}{=}$	is defined as
$E[\,]$	expectation
Ψ	QPC detector proposed in [4, 5]
Υ	QPC detector proposed in this thesis and [6]

Notation convention

For conciseness the notation $B(f_1, f_2)$ is often used in this thesis to denote the discrete bispectrum at the *discrete bifrequency* (k, l) which corresponds *most closely* to the normalised bifrequency (f_1, f_2). If the frequencies f_1, f_2 fall exactly on discrete frequency bins, then $(k = Mf_1, l = Mf_2)$. If they fall in between the discrete bins then $(k \approx Mf_1, l \approx Mf_2)$. A similar convention is also used for the squared bicoherence function $b^2(k, l)$.

Acronyms and Abbreviations

AR	linear auto-regressive model
ARMA	linear auto-regressive moving-average model
AWGN	additive white Gaussian noise
CP	constant phase realisation (as opposed to PR signal)
DFT	discrete Fourier transform
dof	degrees of freedom
FFT	fast Fourier transform
HOC	higher order cumulants
HOS	higher order statistics
iid	independent identically distributed
IT	inner triangle of discrete bispectrum
LTI	linear time invariant
LPC	linear prediction coding
Lx	laryngograph signal
MA	linear moving-average model
OT	outer triangle of discrete bispectrum
PD	principal domain
pdf	probability density function
PR	phase randomised signal (as opposed to CP signal)
PS	pitch-synchronous
QPC	quadratic phase coupling
SNR	signal to noise ratio (in dB)
SOS	second order statistics
UC	uncoupled (as opposed to QPC)
VT	vocal tract
WGN	white Gaussian noise
WSS	wide sense stationary

Chapter 1
Introduction

This thesis is concerned with the application of techniques based on higher order statistics (HOS) to speech signals, concentrating on one HOS measure, the bispectrum, in particular. The bispectrum is a third-order frequency-domain measure, which contains information that conventional spectral analysis techniques cannot provide. Along with other HOS measures, the bispectrum has been developed as a signal processing tool over a period of 30 years, and has been used to provide information about signals which are non-Gaussian, non-stationary and/or nonlinear[1] in a wide variety of signal processing applications, including economic time-series, underwater signal processing, machine condition monitoring and speech.

Of primary interest in this thesis is the fact that the bispectrum can reveal information about nonlinear signal generation mechanisms, in particular mechanisms that contain quadratic nonlinearities. The identification of the signatures of such nonlinearities can be an important step toward the development of new nonlinear models for signals. This is particularly interesting for speech signals, since speech processing is one field which is dominated by linear models, despite the physical arguments, and some experimental results, which indicate that such models are not very suitable models.

There has been interest in using HOS techniques for speech processing since the mid-1980s, but most of the proposed applications to date have been concerned with estimation of conventional speech model parameters in noisy environments. Much of the work in this area has rested on assumptions that speech has certain HOS properties and noise has certain other HOS properties, without ever verifying that this is indeed the case. Consequently, it is very difficult to identify the reasons behind the mediocre results which have often been reported (for example in [7]) for such applications.

The approach taken in this thesis is different: the aim is to lay the foundations of an understanding of what the bispectral properties of speech signals are, since only when

[1] In the sense that they are produced by nonlinear systems.

these properties are fully understood can they be sensibly exploited. The emphasis is thus on speech *analysis* rather than any particular application. It is hoped that the results of this work will then prove useful to others who wish to use HOS for speech processing applications, since it will have helped to establish where, if anywhere, the interesting properties occur.

This aim is achieved in two steps. Firstly an understanding is developed of what new information the bispectrum of a speech signal provides. This involves; a detailed assessment of the properties of bispectrum estimators, to make sure that they can be used with some degree of confidence; an assessment of the performance of existing bispectrum-based quadratic nonlinearity detectors in practical scenarios, and; the development of robust estimation algorithms which can reliably estimate speech bispectra even in noisy backgrounds.

Secondly, these new signal processing tools are applied to a large speech database, since this overcomes some of the limitations of previous work on small data sets. The questions which this analysis attempts to answer are; What are the bispectral properties of speech signals ? ; Does speech exhibit significant quadratic nonlinearities ? ; How do the bispectral measures vary from speaker to speaker, from sound to sound, and from utterance to utterance ? How are different HOS measures of speech signals related to one another ? Can any of this information be utilised in future nonlinear speech models, or robust speech processing systems ?

This work contributes several new results to the field of bispectral analysis which may prove useful to other researchers in fields other than speech processing. These include a number of new results concerning the estimation of the bicoherence (a measure related to the bispectrum) of harmonic signals in noise; the development of a new quadratic nonlinearity detector which overcomes ambiguity problems which occur with existing detectors; and the development of a new robust bicoherence estimation algorithm which is more resistant to transient noise contamination than existing estimators. In addition there are several speech-specific contributions, including a consideration of the issues surrounding pitch-synchronous analysis for bispectrum estimation, and, most importantly, the detailed bispectral analysis of a specially constructed speech database.

1.1 Structure of Thesis

The remainder of this thesis is organised as follows; Chapter 2 gives an overview of the ideas which underly speech processing, and discusses in detail the reasons why the use of the linear model for speech production might be questioned. This serves to set the scene for the introduction and development to the HOS measures used in this thesis, which forms Chapter 3. Starting from a detailed mathematical definition, the assumptions on which the HOS measures are defined are relaxed to arrive at usable measures. The properties of estimators of the bicoherence are considered in some detail.

The next two chapters are concerned with the practical use of bicoherence measures for the detection of nonlinearities in real signals. Chapter 4 describes how problems can occur in the interpretation of bicoherence plots unless certain (strong) assumptions hold. Given that these assumptions often do not hold in reality, a new QPC detector is developed, and its properties are investigated. Chapter 5 describes how the bicoherence estimator can be susceptible to transient noise contamination, and develops an estimation algorithm for reducing the effect of such transients.

Chapter 6 describes the particular problems which are encountered when applying bicoherence analysis to speech signals, and sets out the techniques which in Chapter 7 are used in the analysis of a large database of specially-recorded speech sounds. The discussion of each set of results follows the results immediately, since this allows greater clarity in the interpretation of the results for different speech types.

Finally, Chapter 8 gives an overview of the main achievements of the work described in the whole thesis, outlines areas where more work is needed, and draws conclusions.

The Appendices contain some detailed calculations, experimental details, plus some analysis and a few results which do not fit centrally into the theme of the thesis.

Chapter 2

Beyond the Linear Model for Speech

2.1 Introduction

In this chapter the physical processes which occur when speech is produced will be examined, followed by an illustration of how most popular speech processing schemes are founded on assumptions that the production process is linear. That this is a gross approximation will be discussed using physical arguments, as well as referencing experimental evidence from literature. The aim of this chapter is thus to set the scene, by describing the shortcomings of existing methods, for the development of the HOS-based speech analysis scheme which will be developed in subsequent chapters.

2.2 Acoustics of Speech Production

Figure 2.1 shows a cross section of the human speech production apparatus. All speech sounds are produced by the coordination of the following processes: air is expelled from the lungs and the flow of air is modulated by the configuration of the *vocal cords*, the *vocal tract* and/or the *teeth* and *lips*. The sophisticated coordination of these components gives rise to the wide range of sounds of which we are capable. The vocal cords are muscular folds which are open during normal breathing but can oscillate during speech, and the space between them is called the *glottis*. The cavities above the glottis constitute the *vocal tract (VT)* , and under muscular control the shape of this tract can be changed to change the sound uttered. The main components of the speech production machinery [1], many of which are shown in Figure 2.1, are the *lungs*, *glottis*, *tongue*, *lips*, and *oral* and *nasal cavities*.

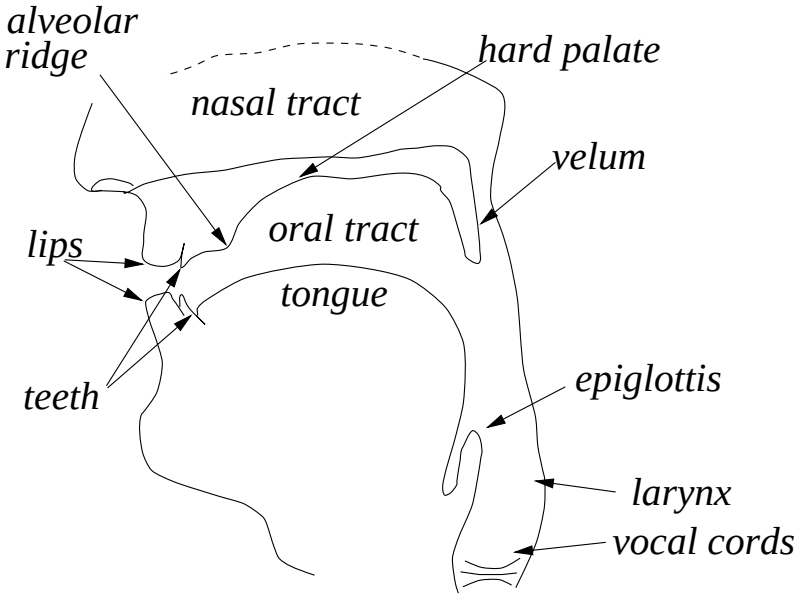

Figure 2.1: Schematic cross-section of the speech production apparatus (based on diagrams from [1, 2]).

2.2.1 Terminology

To be able to describe the speech production process in more detail, it is helpful to divide speech sounds into a variety of classes [1], according to the following criteria;

- Is the VT configuration steady during the production of the sound ? Sounds in which it is are termed *continuants* whilst those in which it is not are termed *non-continuants* [8].
- Are the vocal folds vibrating ? Sounds in which they are are termed *voiced* (or sonorant) as opposed to those in which they are not which are termed *unvoiced* (or voiceless). This situation is sometimes also described by saying that the glottis is vibrating.
- What is the place of articulation ? This is the place at which the VT is narrowest. For example for the sound at the end of the word "ham" the lips are closed, and this sound is termed *bilabial*. If the constriction is progressively moved further back down the VT the sounds are termed *labiodental*, *dental*, *alveolar*, *retroflex*, *palato-alveolar*, *palatal* and *velar*.
- What is the manner of articulation ? This question distinguishes between sounds in which the airstream is completely blocked (e.g. by the lips) which are called *stops* to those in which it is partially blocked in such a way as to cause turbulence

(*fricatives*).

- Is the velum raised ? The velum, a flap of muscle at the rear of the VT, can be lowered to couple the nasal cavity with the oral cavity (for *nasal* sounds), or raised to block off the nasal cavity (for *non-nasals*).

This is a simplistic overview, with some omissions, of a sophisticated system for describing speech sounds which is described in more detail elsewhere [1, 8]. It is presented to provide terms which will be very useful in the description of how speech sounds are formed. In this thesis the focus is on continuant sounds, such as vowels, nasals and fricatives, and so the processes involved in their production will now be described in more detail.

2.2.2 Voiced Sounds

The sound source for voiced sounds is the vibration of the glottis. This oscillation is caused by a cycle of events [9, p60][8, p112]: First, the diaphragm rises, compressing the lungs and forcing a stream of air upward toward the VT. If the glottis is closed then the pressure below it will build up until the pressure forces the vocal cords apart so that the air can escape into the VT. When the forces pushing the folds apart (due to the air flow) equal those pulling them together (due to their elasticity) the process reverses, and the vocal cords begin to close again. In addition, the pressure difference above and below the glottis gives rise to Bernoulli forces[1] which pull the vocal folds together even faster. Consequently the vocal cords snap shut rapidly, and the cycle recommences.

The snapping shut of the glottis can be thought of as some sort of impulse-like excitation to the speech production system. The rate of vibration of the glottis is controllable by the speaker and determines the *fundamental frequency* f_0 of the speech, which is experienced by a listener as the *pitch* of the speaker's voice. The other properties of voiced sounds which, for example, distinguish between the words "heed" and "had" are determined by the configuration of the VT, and primarily by the place of articulation, as described in Section 2.2.1 above. In addition, the VT configuration affects the vocal cord vibration, since it determines the acoustic impedance which the vibrating glottis

[1] Bernoulli forces occur if there is a difference in fluid pressure between opposing sides of an object [8].

"sees".

2.2.3 Nasals

For nasal sounds the velum is lowered, allowing the nasal cavity to couple to the oral cavity. In the English language, nasals occur only when the oral cavity is completely blocked, and so all outward air flow occurs through the nose. Because of this fact, the configuration of the oral tract has only a small effect on the acoustics of nasal sounds.

2.2.4 Fricatives

For fricative sounds the close proximity of one articulator to another results in the air flow becoming turbulent. This can occur in a variety of places, but all such sounds are characterised by the large quantity of random-like noise which they contain. Fricatives can be voiced or unvoiced.

2.3 Modelling Speech

The obvious complexity of the speech production process has not deterred researchers from trying to construct simple models of speech production. There have been many attempts to understand the mechanisms behind speech production and to produce speech by artificial means ([10] provides a detailed review). These date back to the efforts of Wolfgang von Kempelen in the 18th Century to produce a "speaking machine" [10], and include some bizarre yet surprisingly succesful experiments on speech sounds [11], early electronic speech synthesizers in the 1930s, as well as more contemporary speech synthesis techniques[12]. The last 30 years have witnessed massive investment into speech processing research, but with only limited success. In this section a description is given of the most widely-used model of speech production, and some problems associated with it. Recent evidence is then cited which seems to indicate that for speech this model is a fundamentally poor one.

2.3.1 Acoustic Models

Since the physical mechanisms involved in speech production are either extremely complicated or inadequately understood, most speech modelling techniques make several simplifying assumptions about the speech production process. These assumptions include :

The VT and the speech sound source are uncoupled; During the production of voiced sounds (i.e. vowels, nasals) each pitch cycle can be divided into two distinct phases[2] - the closed phase (i.e. when the glottis is closed) and the open phase. It is common to assume that the VT is a cavity closed at the glottis end, but this is evidently only true during the closed phase, since during the open phase the sub-glottal volume of the lungs can also have an influence [8, p163][13].

For unvoiced sounds the situation is even worse since the turbulence can be produced at a variety of places in the VT (for example the lips for the end of the word "life"). In such a case is it really valid to say that the properties of the turbulence are totally independent of the VT configuration ? [8]

Effects due to the non-rigid VT walls, heat conduction and viscous loss; These are usually ignored or modelled in a very crude way since there is no rigorous way of dealing with them [8, p186].

Acoustic Waves propagate as plane waves; By making this assumption, simple acoustic equations can be used to determine the resonance frequencies of the VT. At high frequencies, when the acoustic wavelength is short compared to the size of the VT cross-section, this assumption is no longer valid.

2.3.2 The Linear Model

Even after the above assumptions have been made, acoustic speech modelling is still an extremely complicated problem. Rather than attempting to deal with this acoustics problem, most speech processing techniques employ an extremely simple model of the speech signal, as shown in Figure 2.2. This model forms the basis of several architectures of speech synthesizer (e.g. the formant synthesizer [10]), as well as forming

[2] Phase here is used to mean "time intervals".

the cornerstone of the most successful speech processing technique - Linear Predictive Coding (LPC)[3].

Figure 2.2: Linear model of speech production.

Techniques based on the linear model dominate all speech processing areas: As well as the obvious link with speech production models, speech synthesis, and speech coding, speech recognition too is dominated by the linear source-filter model of speech production[4]. Even the speech processing technique of cepstral analysis (an ostensibly nonlinear technique for vocal tract deconvolution) is fundamentally based on the linear model of speech production.

Many of the properties of the linear speech model, and the assumptions which underpin it, can be related to the acoustic assumptions described above. In addition, techniques based on the linear model make the following assumptions:

Phase is not Important; LPC-modelled speech has the same power spectrum as the original speech, but it does not necessarily have the same phase. The assumption that the resynthesized speech does not need to have the same phase as the original signal underlies the LPC approach, and in this sense LPC ignores the importance of the speech signal phase. Whether or not phase is a perceptually important property of speech signals is a long-running question: In 1877 Helmholtz observed "The quality of the musical portion of a compound tone depends solely on the number and relative strengths of its partial tones and in no respect on their differences of phase " [14, p127]. This view has led many researchers to view the matter as closed, and subsequently it is often assumed [15, 8] that phase plays no role in speech communication. However, although phase does seem to play only a small role in speech *intelligibility* it has been found to play a much more important role in speech *naturalness* [9], and in general signal *quality*, evidence

[3] Although it is called "Coding" it is in fact also used extensively in synthesis [2, p75] and recognition [2, p109] systems.

[4] Although Hidden Markov-Model and Neural Net architectures are used to deal with nonstationarity of the signal (i.e. in the time domain), the signal is still viewed as a concatenation of states which can be adequately described using linear based features (e.g. LPC parameters).

of which comes from experiments on musical timbre[5] [16, 17]. The fact that LPC ignores the signal phase thus appears to be an indication of weakness as far as speech naturalness is concerned.

The excitation of the linear filter is Gaussian; The parameters of the linear speech model used in LPC techniques are determined from the second-order statistics (SOS) of the speech signal (i.e. the autocorrelation/autocovariance function and the power spectrum). These statistics only *fully* characterise a signal's pdf if the signal is Gaussian (since a Gaussian pdf is fully described by just its mean μ and variance σ^2), and if speech is non-Gaussian then the SOS only encapsulate *part* of the available information. Thus it follows that if speech is indeed non-Gaussian, then the conventional LPC approach fails to use the potentially useful higher-order information. Although previously speech has often been modelled as having a Gaussian pdf [18, p32], recent evidence [19, 20, 21] appears to indicate that in general it has a non-Gaussian pdf.

The filter is linear; The AR model on which the LPC technique is based is known to have a weakness concerning the representation of sounds such as nasals whose spectra contain zeros [10, p113][8, p269]. Although this is a weakness of the AR linear model, it does not *necessarily* indicate that speech is nonlinear. However, other evidence, which will be discussed in Section 2.3.3 below, has cast doubt on the suitability of this filter model for speech production.

Although the applicability of each of these assumptions to speech analysis can be questioned, the ease of implementation of this model has led to its widespread use in many speech processing fields. Furthermore, judging by the criterion of speech *intelligibility* [2, p100] linear-based models perform well [9]. From one viewpoint, the suitability of the assumptions on which the linear model is based has been measured by the fact that linear-based techniques result in intelligible speech.

Recently however, interest has turned to different criteria of speech assessment - those of *quality* and *naturalness* [8, p57ff]. Although linear-based models are capable of producing highly intelligible speech, often the speech does not sound very *natural* [10, p118]. This observation applies to the simplest linear-based techniques rather than to the more sophisticated ones (such as CELP), but it raises the question as to whether simple nonlinear speech models may be better at capturing some of the naturalness in

[5] "Timbre" is tone colour, an objective measure similar to quality.

real speech.

This problem has in turn led to the investigation of alternative speech production models, and to new signal processing tools which encapsulate the information which is in human speech signals but which linear-based techniques (such as LPC) do not use.

2.3.3 Evidence for nonlinearities in Speech

In two much cited papers, Teager [22] and Teager and Teager[23] describe experiments using hot-wire anemometers which suggest that the linear model of speech production has serious deficiencies. Central to their findings are that the relationship between the air flow and the pressure in the vocal tract is not constant (whereas the linear acoustic assumption dictates that it should be), and that the actual flows observed in the vocal tract are much larger than those precited by passive acoustic theory.

Subsequently there has been a steady growth of interest in the development of speech analysis/synthesis techniques based on nonlinear models, which can encapsulate signal phase information, cater for non-Gaussian signals, deal with the rapidly changing speech dynamics, and/or can reveal information about nonlinear signal production mechanisms. A comprehensive review of these techniques which exist for these tasks was presented recently [24].

These techniques can be divided into two types:

1. Those that use a nonlinear dynamics (and chaos theory) approach to try to model the rapidly-changing dynamics of the speech [25]. The nonlinear dynamics perspective is closely related to waveform modelling approaches, but although it has appeal for such reasons, the link between nonlinear dynamics representations of speech and the physics of speech production is obscure. Speech modelling using nonlinear dynamics requires an analysis framework (such as those discussed by [24, 26]) very different to the conventional speech analysis framework. This makes it difficult to gauge what extra information the nonlinear models provide.

2. Those that still treat the speech signal as a concatenation of quasi-stationary states (in the same way as LPC models, and HMMs do), but to use nonlinear models to describe these states. These techniques are attractive because they plug neatly into the currently-existing framework of speech analysis, but with

new speech filters which are more sophisticated than simple linear filters. These techniques suffer from the same problems, such as poor representation of coarticulation, that beset conventional frame-based speech analysis suffers, but they are conceptually straightforward, and can provide useful information about speech production. The interest in this area has included nonlinear prediction models [27, 28, 29], Volterra models of speech signals [28, 29] as well as the detection and characterisation of nonlinearities in speech signals. It is in this last area that the current work is focussed.

2.4 Higher Order Statistics

HOS techniques were revived in the 1960s by statisticians with a view to extracting new information from existing data. Initial interest in the fields of economics [30] and geophysics [31] expanded over time as researchers in other fields became interested in HOS techniques. Subsequently, HOS techniques have been used for signal analysis in fields as diverse as underwater acoustics [32], machinery noise [33], plasma physics [34], optics [35] as well as speech processing [36]. The main reason for this interest in HOS techniques is that they can provide information which supplements the information available from traditional SOS measures (the autocorrelation function and the spectrum).

Since HOS techniques can be regarded as *extensions* of existing SOS techniques to higher orders, practically any technique which involves SOS can be reformulated in terms of HOS. The motivation for doing this is that HOS measures are, at least in theory, "blind" to Gaussian noise [37], and so it may be possible to achieve cleaner estimates in noisy environments using the HOS formulations. Whether or not this is a wise thing to do is not always clear, for HOS techniques are generally more complicated, more computationally expensive, and more difficult to use[6] than their SOS counterparts. Nevertheless a large portion of the completed work in HOS applications has been concerned with trying to estimate "old" quantities in "new" ways. As these techniques *exploit* the difference in HOS properties between speech signals and noise, it will be convenient to call these techniques "Exploitative techniques".

In addition, HOS techniques can be used to reveal information about nonlinearities

[6] In the sense that the HOS estimates generally have higher variances than their SOS counterparts.

which is *simply not available* using conventional techniques. These techniques do not connect to the existing signal processing frameworks in such a simple way, and perhaps because of this reason, there has been less interest in this area. However, it is in this area that the current work is focussed. In particular the current work deals with third-order frequency-domain polyspectral measures - the *bispectrum* and *bicoherence*. Since these techniques provide new information which simply cannot be obtained by conventional methods, these types of techniques will be called "Exploratory techniques".

2.4.1 Previous Work

It might appear sensible to first use exploratory techniques extensively, in order that the HOS properties of speech are well described, and their origins are well understood, before going on to develop exploitative techniques which use the knowledge gained from the exploratory methods. However, this has not been the case for HOS and speech analysis, and the exploitative techniques, which are often based on somewhat flimsy evidence of the HOS properties of speech signals, have dominated. Table 2.1 summarizes the main subject areas of speech analysis for which HOS techniques have been employed.

	Exploitative			Exploratory		
	T	T/F	F	T	T/F	F
AR model	[38, 39, 40, 41]					
ARMA model				[42] [43]		
Recognition	[7]					
Characterisation						[44, 45] [36][46]
Reconstruction / Enhancement	[47]	[40, 41] [48]			[49] [50]	
Event detection	[51, 52] [53]			[21, 20]	[54, 55]	[36]
Pitch	[56]					

Table 2.1: References to previous work on HOS and speech. T=time domain, T/F=time and frequency domain, F=frequency domain method. [Square brackets] are used to enclose references to work from one research group.

It is pertinent to mention here that although there has been a great deal of academic interest in speech processing with HOS, to date there has not been a single full journal paper published on the subject. The reasons for this may be attributable to the

technical difficulties of applying HOS techniques to speech signals, and perhaps also to the practical difficulty of researchers being able to simultaneously span two rapidly expanding research fields, such that it is extremely difficult to remain at the cutting edge of both HOS and speech. In the following sections a survey is presented of the existing literature, mostly from conference proceedings, in this field. Because of the nature of conference papers, many of the findings described below are only "preliminary", which at best means inconclusive, and at worst means incomplete.

2.4.1.1 Exploitative Techniques

Many of the exploitative techniques were pursued as a result of an opinion widely held during the late 1980's and early 1990's, that SNRs could be improved by carrying out computations in HOS domains [37]. This idea was based purely on theoretical results, and it was only later that it became apparent that for finite data lengths there was a price to be paid for the noise-robust properties of the HOS techniques - namely that the variances of the HOS estimates required were in general much higher than the variances of the traditional SOS estimates.

The (perhaps naive) idea that speech properties would somehow "shine through" Gaussian noise in the HOS domain provided motivation for many researchers in the field. Cumulants, the time-domain HOS measures akin to the autocorrelation function, were used for pitch period estimation[56, 57], speech endpoint detection [51, 52], voiced/unvoiced decision [53] and for the determination of LPC parameters [39, 7, 40, 41, 47].

Most results reported were preliminary, and so it is difficult to comment on the results. However, in Speech Recognition experiments [7] it was found that the HOS techniques performed worse than traditional methods at high SNRs, an indicator of the estimation problems which have plagued HOS techniques from the outset. Furthermore in a HOS-based LPC application [47] it was found that HOS methods gave at best no improvement over existing techniques.

However, some interesting theoretical relations were discovered [58] which made a link between residual-weighted LPC techniques (previously regarded as a somewhat ad-hoc modification of LPC) and fourth-order cumulant slices. This indicated that the residual-weighted LPC technique could be viewed as a HOS technique.

Frequency-domain measures (based on the bispectrum and bicoherence) were also put

to exploitative ends. In a preliminary study [48] it was proposed that speech could be enhanced by first estimating its bispectrum and then reconstructing the speech from its bispectrum, the idea being that, since the speech is non-Gaussian and the background noise is Gaussian, the bispectrum will be sensitive only to the speech, and not the noise. Again no account was taken of the fact that HOS estimators have high variances, and probably as a consequence of this, only a small improvement in SNR was reported, despite greatly increased computational cost.

Test statistics for speech endpoint detection were also developed, based on the bispectrum and bicoherence [55, 54] but few comparisons were made with existing techniques. A problem with many of the HOS measures is that they can be inadvertently sensitive to SOS too, and so sometimes SOS properties (such as correlation and energy) have been mistaken to be HOS quantities. More will be said on this matter in Chapter 3.

2.4.1.2 Exploratory Techniques

In the exploratory field, the ground rules are less clear since the researcher does not know what s/he is looking for. The key motivation in this field is to find out *what* the HOS properties of speech signals are, and to find out *if* they can provide useful information. As a question first raised over ten years ago [36], this has yet to be answered.

The first attempt to use HOS for speech processing involved a description of sustained speech sounds (vowels and fricatives) using the bispectrum and bicoherence [36]. The analysis was carried out using a pitch-synchronous segment-averaging technique, which is the same as the one used in this thesis. It was reported that unvoiced fricatives had approximately zero bispectra, in contrast to voiced sounds which had non-zero bispectra. This paper, which only described preliminary findings, has been repeatedly used by those working on exploitative techniques as "proof" that their approaches are valid. In the author's opinion the evidence is far from conclusive.

Bicoherence analyses of speech sounds for pathology detection have also been reported [44, 45], but the results are again inconclusive. A related application to the analysis of musical signals [46] also failed to identify any key virtues in the HOS approach. A more complete analysis was published recently [28] in which the bicoherences of a few vowel sounds were analysed. Evidence of significant coupling was reported *at* the harmonics of the speech pitch frequency.

Recently it has been shown [28, 29, 6, 4] that the assumptions made in some applications [44, 45, 46] render the bicoherence ambiguous. This problem is elucidated, and a novel solution is proposed, in Chapter 4.

In the time domain there have been relatively few attempts to investigate the HOS properties of speech signals from an exploratory point of view. The characterisation of speech "events" by their higher-order moments (HOMs) has been described [20, 21], based on the idea the extra information provided by the HOMs can aid identification of important parts of speech signals, with the aim of performing automatic segmentation. The results of these studies were again somewhat inconclusive, mainly because of the difficulties of estimating the HOMs from short data lengths.

There have been only two attempts to use HOS techniques in a exploratory parametric framework. The approach which has been considered [42, 43] is the determination of linear ARMA model parameters for speech using HOS (LPC systems use only an AR model of speech.) Because of the extra phase information which HOS measures provide, it is possible to estimate the parameters of non-minimum phase systems using such techniques, in contrast to SOS methods which estimate the minimum phase system with the same spectrum as the unkown system. However, once again the results are somewhat inconclusive.

2.5 Summary

In this chapter the physical process of speech production have been reviewed, and some of the assumptions made by conventional approaches to speech modelling have been described. The weaknesses in these assumptions have been identified, and higher order statistics (HOS) techniques have been introduced as tools with the potential for developing better models of speech production. The previous work on HOS and speech has been reviewed, and it has been established that to date there has been insufficient investigative work carried out on characterisation of speech sounds by their HOS properties. Such work is necessary before these HOS properties can (if at all) be intelligently exploited, and therefore the goal of this study will be to establish *what* the HOS properties of speech signals are, and to find out *how* these properties arise.

Chapter 3

Bispectral Analysis

3.1 Introduction

The motivation for using Higher Order Statistics (HOS) measures in signal processing arises from the useful properties which HOS measures possess. As discussed in Chapter 2 it is helpful to divide the HOS techniques into two groups;

Exploitative Techniques : these aim to provide noise-robust estimates of quantities which are usually determined via second-order techniques.

Exploratory Techniques : these provide information which is simply not available using conventional measures.

The explotative techniques are usually founded on the assumption that the HOS properties of the signal of interest are *different* to the HOS properties of the measurement noise. Therefore the SNR can, in principle, be increased by developing algorithms which make use of these HOS properties [37]. In contrast, the exploratory techniques have been developed from a different perspective - that of providing **new** information which conventional techniques are simply unable to provide.

Historically, many of the HOS techniques were developed by statisticians whose primary goals were the pursuit of new and interesting statistical results, rather than the application of these techniques to real-life situations. Since the exploitative techniques use the same framework as exisiting second-order techniques, much of the engineering interest in HOS, which increased greatly between 1980 and 1995, was concerned with "bolting on" new HOS techniques to existing algorithms. In contrast, the exploratory techniques cannot be interfaced with the existing techniques in such a simple way, and so there was generally less interest in these techniques. As a result, the exploitative techniques received most attention.

However, over time it has become apparent that practical problems, usually to do with the estimation of the HOS measures from finite-length data records, place severe limitations on the usefulness of the exploitative techniques. This appears to have resulted in a shift in recent years towards the exploratory techniques. Progress in research into these techniques can still be useful for the development of exploitative techniques, however, because new information about how signal and noise properties differ can be exploited in new noise-robust algorithms. The current work concentrates on exploratory techniques - i.e. providing information which supplements information provided by conventional techniques.

In this chapter the statistical background of HOS techniques is described, starting from basic mathematical signal assumptions and going, via the definitions of *cumulants*, to *polyspectra*, their frequency-domain analogues. To develop a focus for the chapters which follow the 3rd-order polyspectrum, or *bispectrum*, receives special attention. Estimation issues are addressed, and some of the useful properties of the bispectrum are discussed. The application of the bispectrum for nonlinearity detection, which is a central theme of this thesis, is introduced in this chapter, but is developed in greater length in Chapter 4.

3.2 Higher Order Statistics

The subject of HOS has been developed on a rigorous mathematical framework which requires many assumptions to be made about the signal of interest. The approach taken here is first to introduce the HOS measures in a mathematical fashion (following [59, 60]), and then, with justifications, to move toward a signal processing / engineering viewpoint (such as [61, 34]), from which subsequent chapters will develop new algorithms and interpretations. The papers cited do not necessarily represent the chronology of developments in the field, since some of the later papers [59, 60] are more tutorial in nature.

Consider a real valued discrete time random process $x(n)$[1]. One of the key assumptions which must be made to develop the HOS measures is that the process is (in some sense) stationary. Now in many signal processing techniques the condition of wide

[1] In keeping with the statistical literature, the discussion begins using the terminology of "random processes", but these will be replaced with the engineering terminology of "discrete signal" in a later section.

sense stationarity (WSS) is assumed. This implies [62, p105] that, for any time indices $t_0, .., t_{N-1}$ and any lag τ, that all the joint moments up to order 2 of the process $x(t_0), x(t_1), x(t_2), .., x(t_{N-1})$ exist and *equal* the corresponding joint moments up to order 2 of the shifted process $x(t_0+\tau), x(t_1+\tau), x(t_2+\tau), .., x(t_{N-1}+\tau)$. In other words, the mean and covariance of the process are not changed by an arbitrary time shift. Stronger stationarity assumptions require the same sort of invariance under time shift of higher order moments. For the time being, for generalisation, it will be assumed that the process is strictly stationary, so that *all* its joint moments are unchanged by an arbitrary time shift.

The process $x(n)$ can be characterised in many ways, for example by its amplitude, its energy or its waveform. The probability density function (pdf) of the process provides detailed information about the distribution of the amplitudes of the process, which can be used to characterise the process. A set of quantities which describe the shape of this pdf are the *moments*.

3.2.1 Moments

The first-order moment m_1 of the process $x(n)$ is just its mean μ, and it provides a measure of location of the pdf. The second-order moment is the variance, a measure of the spread of the pdf. Higher-order moments exist too, such as the *skewness* (a measure of asymmetry in the pdf) and the *kurtosis* (a measure of sharpness of peak of the pdf). A function called the moment generating function (MGF) can be used to encapsulate all the moments. The first coefficient in the Taylor expansion of the MGF is the mean, the second the variance, and so on.

In general the kth order moment about the mean[2] can be calculated by taking an expectation over the process multiplied by $k-1$ lagged versions of itself [63]:

$$\begin{aligned} m_2(\tau_1) &\triangleq E[x(n)x(n+\tau_1)] \\ m_3(\tau_1,\tau_2) &\triangleq E[x(n)x(n+\tau_1)x(n+\tau_2)] \\ m_4(\tau_1,\tau_2,\tau_3) &\triangleq E[x(n)x(n+\tau_1)x(n+\tau_2)x(n+\tau_3)] \\ \vdots &\quad \vdots \\ m_k(\tau_1,..,\tau_k-1) &\triangleq E[x(n)x(n+\tau_1)..x(n+\tau_{k-1})] \end{aligned}$$

[2]Thus if the process has nonzero mean, the mean should be subtracted from it first.

The variance σ^2 is then given by $m_2(0)$, the skewness γ_3 by $m_3(0,0)$, and so on.

Now instead of dealing with the *moments* of the process, an alternative set of measures, which have some useful mathematical properties [59], can be used instead. These are the *cumulants*.

3.2.2 Cumulants

Whereas the moments can be defined in terms of the Taylor expansion of the MGF (Moment Generating Function) [3], the cumulants can be defined in terms of the Taylor expansion of the CGF (Cumulant Generating Function). The CGF is just the natural log of the MGF, so cumulants are very closely related to the moments. However, cumulants posess certain properties which lend themselves well to the development of new HOS techniques, and because of these properties most HOS techniques are developed in terms of cumulants and not moments.

Perhaps the most important of these properties are those concerning Gaussian processes. A Gaussian process is completely characterised by its mean and variance only, and it is easy to show (see Appendix A.1) that the first-order cumulant of a Gaussian process is equal to the mean, that the second-order cumulant of a Gaussian process is equal to the variance, and that all higher-order cumulants are identically zero. This property suggests that measurement noise, which is often assumed to be Gaussian, disappears at third- and higher-orders. This raises the possibility that if the process of interest is non-Gaussian, then its properties will "shine through" the noise in the higher-order domains. This remains one of the key motivations for research in HOS methods.

However, since the mathematical properties of cumulants are not central to the current investigation, the reader is referred to [63] for a detailed description of their properties.

The cumulants of $x(n)$ can be calculated by first calculating the moments of the process, and then using the simple relations which exist between moments and cumulants [63] to determine the cumulants. For zero-mean processes ($m_1 = 0$), the second- and third-order cumulants $c_2(\tau_1)$ and $c_3(\tau_1, \tau_2)$ turn out to be *the same* as the second- and third-order moments $m_2(\tau_1)$ and $m_3(\tau_1, \tau_2)$ respectively. In practice it is always easy to subtract any nonzero mean from the data as a preprocessing step, and because the zero-mean case is usually easier to analyse, the following sections will make the zero-mean

assumption.

Now that cumulants and moments have been introduced, the frequency-domain representations which are used as the basic measurement tools in this thesis will be introduced - these are the *Polyspectra*.

3.2.3 Polyspectra

The relation between time-domain and frequency-domain measures forms the foundations of much of modern signal processing. The discrete Fourier Transform, the DFT, provides a means for transforming from time- to frequency-domain, and vice versa. This is useful because signal properties do not always manifest themselves in the signal waveform, and transforming to the frequency-domain can expose periodicities in the measured signal, and can aid understanding of the processes which produced the signal. In this section the familiar second-order time- and frequency-domain measures are extended to give new, higher-order measures. The generalisation of the power spectrum to higher-orders forms the family of polyspectra, one of which, the bispectrum, is at the core of this thesis.

To recap, consideration has been given above to the cumulants and moments of a strictly stationary zero-mean process $x(n)$. For mathematical convenience, it will be assumed that samples of $x(n)$ which are far apart in time are statistically independent. This constitutes a "mixing condition" [64, p8], and means that there are no long-term corrrelations between samples of the process.

Given a further assumption, concerning the summability of the cumulants [62, 63, 64][3], the Discrete Fourier Transform (DFT) of the nth order cumulant can then be taken to give the nth-order *cumulant spectrum* [63, 59] or *polyspectrum* [62]. This is simply a generalisation of the Wiener-Khintchine relation (the well known relation between the second order measures (the autocorrelation function $R(\tau_1) = c_2(\tau_1)$ and the power spectrum [3]) to any order.

Starting with the familiar second-order case, the (discrete) power spectrum $P(k)$, with discrete integer frequency k can be written *either* as the DFT of the autocorrelation function $R(\tau_1) = c_2(\tau_1)$, where τ_1 is a discrete lag, *or* as a product of two DFTs $X()$

[3]This means that the sum of any cumulant sequence over all lags is less than ∞.

whose sum frequency is zero[4]:

$$P(k) \stackrel{\triangle}{=} DFT\left[R(\tau_1)\right] \equiv DFT\left[c_2(\tau_1)\right] \equiv E[X(k)X^*(k)]. \tag{3.1}$$

The power spectrum can be thought of as a frequency decomposition of the autocorrelation function, and these two quantities are related to each other in the following way; the sum of the power spectrum $P(k)$ over all frequencies k is equal to the zero-lag autocorrelation function $R(0)$. This quantity is called the variance σ^2 of the process. The term on the extreme right-hand side of Equation 3.1 will be the one of most interest in this work.

Now extending this idea to higher-orders, the 3rd (and 4th) order *polyspectrum* can be written as the double (triple) DFT of the 3rd (4th) order cumulant sequence:

$$\begin{aligned} B(k,l) &\stackrel{\triangle}{=} DFT^2\left[c_3(\tau_1,\tau_2)\right] \equiv E[X(k)X(l)X^*(k+l)] \\ T(k,l,m) &\stackrel{\triangle}{=} DFT^3\left[c_4(\tau_1,\tau_2,\tau_3)\right] \equiv E[X(k)X(l)X(m)X^*(k+l+m)], \end{aligned} \tag{3.2}$$

in which the DFT^2 indicates that $B(k,l)$ is the double-DFT of the third order cumulant sequence.

These equations link time-domain measures (the cumulants) with frequency-domain measures (the polyspectra), and the quantities $B(k,l)$ and $T(k,l,m)$ are known as the *bispectrum* and *trispectrum* respectively, and form the two most widely studied higher-order spectra (primarily because practical constraints all but rule out consideration of the higher orders). Note how, in a similar way to the power spectrum, each polyspectrum has frequency indices which sum to zero, although for the bispectrum and trispectrum several *different* frequency components (e.g. k, l *and* $k+l$ for the bispectrum) contribute to each polyspectrum value. This thesis will focus on the third-order polyspectrum, the bispectrum, and the interested reader is referred elsewhere for matters concerning trispectral analysis [65, 66].

3.3 The Bispectrum

The bispectrum can be thought of as a frequency decomposition of the third-order cumulant, and it follows that the process skewness γ_3, which is the zero-lag cumulant

[4] Using the fact that $X(-k) = X^*(k)$.

$c_3(0,0)$ is equal to the bispectrum summed over all frequencies. This is similar to the way in which the variance of a process is related to its power spectrum and second-order cumulant (or autocorrelation function).

The bispectrum $B(k,l)$ (Equation 3.2) is a complex quantity, and so it has magnitude $|B(k,l)|$ and *biphase* $\angle B(k,l)$. Both magnitude and phase have two independent frequency axes, k and l, to locate the bispectral content at the *bifrequency* (k,l) and so each is a 3-d quantity. Figure 3.1 illustrates how the DFT at the three frequencies k, l and $k+l$ contributes to the bispectral content at bifrequency (k,l). It is the fact that the bispectrum measures[5] the interaction *between* frequencies that lies at the heart of its usefulness.

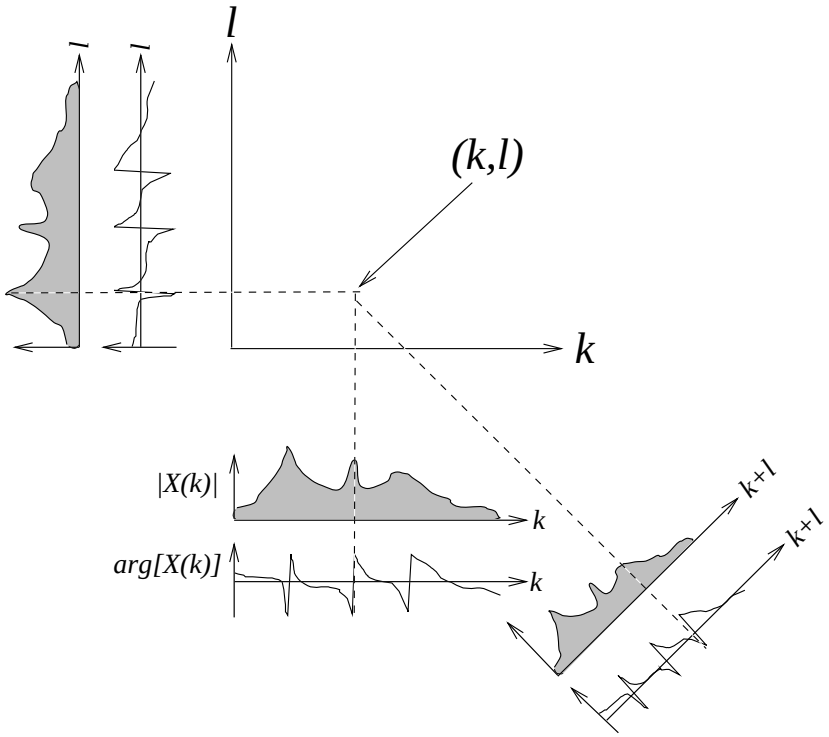

Figure 3.1: The origin of bispectral content - the bispectrum $B(k,l)$ contains contributions from the magnitude and phase of the DFT at the three frequencies k, l and $k+l$.

3.3.1 The Principal Domain

In a way similar to that in which the discrete power spectrum has a point of symmetry at the folding frequency $f_s/2$, the discrete bispectrum has many symmetries in the k,l

[5] In a way to be discussed at length in Chapter 4.

plane [67, 63, 68][6]. Because of these, it is only necessary to calculate the bispctrum in the non-redundant region or *principal domain* (PD) as shown in Figure 3.2. The PD can be divided into two triangular regions in which the discrete bispectrum has different properties, the *inner triangle* (IT) and the *outer triangle* (OT). Further discussion of these properties will be given in Section 3.6, but it is worth noting here that in the current work the IT is of primary interest, and so bispectral plots in this thesis will usually only show this triangular region.

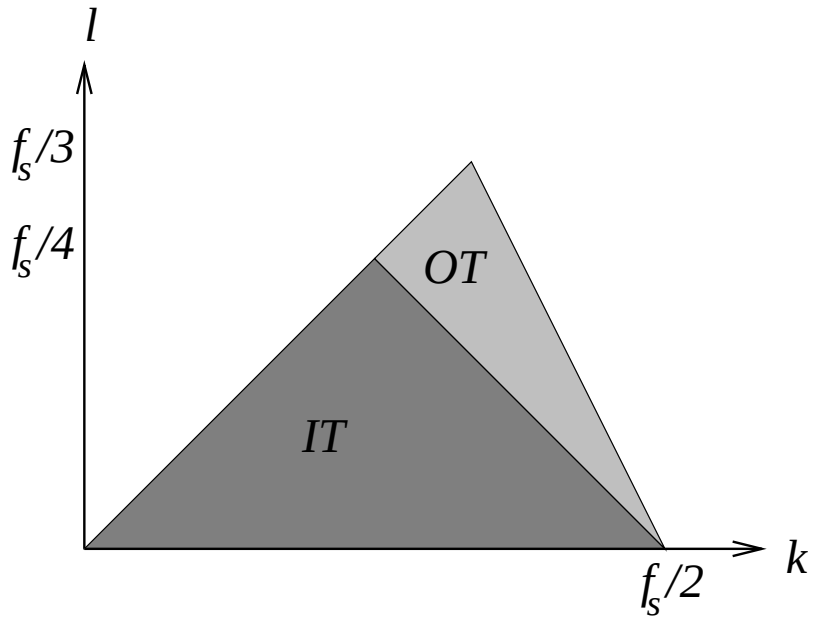

Figure 3.2: The principal domain of the discrete bispectrum.

3.3.2 Establishing an Engineering Framework

The interpretation of bispectral plots is somewhat dependent on the type of process under consideration; from the way in which the bispectrum has been introduced it is clear that it has a close relationship with the process skewness (since $B(k,l) = DFT[c_3(\tau_1,\tau_2)]$ and $c_3(0,0) = \gamma_3$ from Section 3.2.1), and this property will be investigated later in this chapter. In addition, the bispectrum can also reveal interesting phase relations between frequency components, which can be used to detect certain types of nonlinearities. However, before these bispectral properties can be discussed in more detail, a practical framework of bispectral estimation must be established. As a first step in this direction, some of the mathematical assumptions used in defining the bispectrum will be relaxed to give a less rigorous but more useful measure. A change in

[6]The continuous bispectrum does not share all these symmetries [67].

terminology is made here, from the statistical concept of the *process* to the engineering concept of the *signal*.

3.3.2.1 Relax assumption that $x(n)$ is stochastic

The framework for polyspectral analysis has been developed with attention to stochastic processes, but, as it will be shown below, it is also interesting to investigate signals with strong deterministic components, such as sinusoids. Such signals have properties which violate the assumptions from which the bispectral results are derived, but it is often possible to find a way around such problems so that useful, usable measurements can be made.

3.3.2.2 Relax assumption that $x(n)$ is stationary

The assumption has been made in Section 3.2 that the signal $x(n)$ is strictly stationary. In fact the level of stationarity required is determined by the polyspectrum order, and in fact the signal need only be stationary to third-order for the bispectrum definition to hold [60].

However, if the signal is deterministic, such as a pure sinusoid with constant amplitude, frequency and phase, then it can no longer be modelled as a stationary process, since its mean and autocorrelation function are time-varying [69]. To appreciate this somewhat subtle point, consider the following argument.

To be stationary, a signal must have a mean which is not time varying. The mean of a random process is the average value over an infinite number of realisations of the process. For a sinusoidal signal with constant parameters, every realisation is identical, and so the mean depends on the time at which the ensemble of realisations are measured. Thus the mean is time-varying, and the process is non-stationary.

Sometimes a sinusoidal signal can be assumed stationary if the sinusoid phase is assumed to be a random variable [64, Section 2.10][69], but this can be an unrealistic assumption for some real-life signals. Despite these arguments, the bispectra of sinusoidal signals can still provide useful information about nonlinearities, and the nonlinearity-detection interpretation of bispectra, which will be developed in Chapter 4, is central to this thesis.

An alternative accommodation of sinusoids into a stochastic framework can be achieved by viewing the sinusoid as a zero variance process with time varying mean [70], but this renders the process highly non-stationary, and so is not useful in the current investigation.

3.3.2.3 Relax assumption that $x(n)$ has summable cumulants

Periodic signals are power signals - they have infinite energy (since they go on forever) but finite power. Consequently the autocorrelation function and the cumulants have infinite sums (which violates the assumption of summable cumulants mentioned above). Furthermore, samples which are well separated in time are no longer independent, and so the mixing condition assumption is also violated. However, by taking a practical viewpoint, the use of data windows for DFT computation (effectively setting the signal amplitude to zero outside the window), has the effect of ensuring that the cumulants *are* summable. As a result of this, Equation 3.2 can still be used. (The choice of data windows will be dealt with in detail in Section 3.7.1 below).

3.4 Estimation

The estimation of the bispectrum is a fascinating subject area which has received much attention over recent years. The underlying approaches are extensions of the well-established second-order power spectrum estimation approaches, but the properties of the bispectral estimates turn out to be rather different to the power spectral estimates. Overcoming the difficulties presented by these properties leads to the development of new measures, which remain closely related to the bispectrum. In this section an overview is given of bispectral estimators, and results are cited from the literature which illustrate the estimator properties, and suggest modifications to the bispectral measures.

3.4.1 Two signal models of interest

Two different signal models have been used in the literature for the development of HOS techniques. These are described below, and will be referred to in following chapters as the M1 and M2 models.

Signal M1 is a stochastic signal model, and M2 is a mixture of deterministic and stochastic signals. It is of considerable interest to develop a framework for estimation and analysis that is independent of the signal model chosen, since in practice the distinction between signal types is often blurred. For example, the ways in which DFT analysers are used for practical spectrum estimation in the laboratory are the same for all signal types, but the interpretation of the results of such spectra depend on the skill of the experimenter. The development of a similar framework for bispectral analysis will involve the consideration of the properties of bispectra of signals conforming to each model, and careful interpretation of the results.

The properties of bispectral estimators have previously been investigated for two separate models;

M1 A causal, linear MA filter $h(n)$ driven by a stochastic process $e(n)$ as shown in Figure 3.3.

$$x(n) = \sum_{m=0}^{\infty} h(m)e(n-m) \tag{3.3}$$

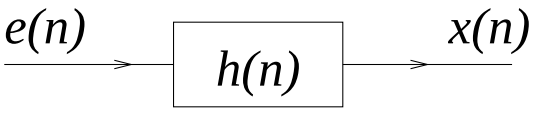

Figure 3.3: Simple linear MA process generating $x(n)$.

Signals conforming to this model are interesting from a HOS perspective only if the input process $e(n)$ is non-Gaussian. This is because a Gaussian $e(n)$ results in a Gaussian $x(n)$ which, as it has already been discussed, has zero cumulants, and hence zero bispectrum (Appendix A.1).

M2 A summation of P sinusoids (deterministic) in additive noise (stochastic):

$$x(n) = \sum_{p=1}^{P} A_p \cos(2\pi f_p n + \phi_p) + v(n) \tag{3.4}$$

where the pth sinusoid has amplitude A_p, normalised frequency f_p and phase ϕ_p. By careful choice of the frequencies f_p and phases ϕ_p of this model, it can generate signals which contain the frequency-coupling and phase-coupling signatures of signals generated by quadratically nonlinear systems. From this linear equation, the signatures of quadratic nonlinearity (i.e. the frequency- and phase-coupling) can be easily switched on or off, as required. The bispectra of signals conform-

ing to this model can reveal these signatures, which are hidden from second order measures, and thus the bispectrum can be used to test for certain types of nonlinearity. This is discussed in more detail in Section 3.6.4 and in depth in Chapter 4.

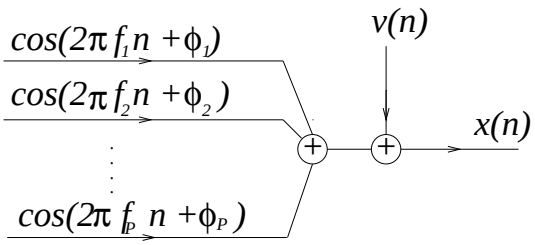

Figure 3.4: Simple harmonic signal model.

3.4.2 Ensemble or Segment Averaging

In [63] it is stated that the bispectra of deterministic signals can be estimated from a single realisation without averaging. This is only true for signals in which measurement noise is zero, a situation which is rarely, if ever, encountered in practice. This being the case, some form of averaging is required to obtain estimates with statistical stability whether model M1 or M2 is appropriate.

Averaging can be carried out in one of two ways;

- Multiple realisations of the process can be obtained, and estimates averaged over these realisations - this is *ensemble* averaging.
- Only one realisation is used, and it is divided up in some way (either in the time or frequency domain) and averaged accordingly.

It is often the case in practice that only a single realisation of a process is available. For example in speech analysis, there is no guarantee that multiple utterances of any one word will actually be realisations of the same underlying process because the speaker does not have perfect control over the speech production system. It thus makes sense to try to use estimates which can be determined from a single realisation of the process.

This approach is only meaningful if the measures concerned are *ergodic*, that is if the statistical properties obtained from a single record are the same as those of an ensemble average. It will be shown in Chapter 4 that bispectral measures can be non-ergodic for

some types of M2 signals, but that even in that case some useful information can be retrieved from the bispectrum. For the time being though it will be assumed that the measures considered are ergodic.

3.4.3 Direct and Indirect Methods

Given a signal $x(n)$, there are two popular techniques for estimating its bispectrum [63]; the *direct* and *indirect* methods. Both have similar computational requirements [71], but in the current study only the direct method is used. This is because it has been found, in the current work, to be easier to implement and conceptually simpler than the indirect method. Both methods give similar results [72].

The direct method is an extension of the Welch periodogram averaging technique for spectral estimation. The procedure is shown schematically in Figure 3.5 and consists of the following steps;

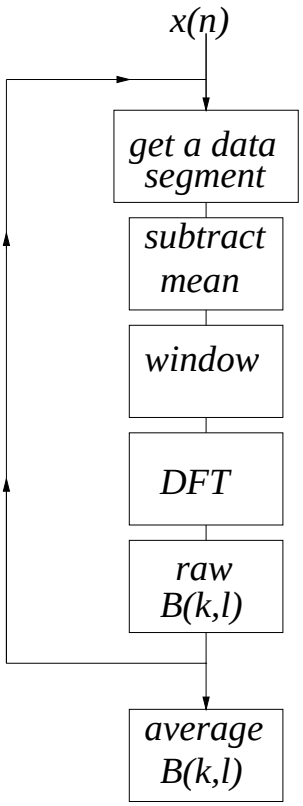

Figure 3.5: Schematic diagram illustrating direct method of bispectrum estimation.

- The signal $x(n)$ $n = 0, .., N-1$ is divided into K segments $i = 0, .., K-1$, each of length M, as shown in Figure 3.5. These segments can overlap, so $K \geq N/M$.

Let the ith segment of $x(n)$ be $x_i(n)$ $n = 0, .., M-1$.

- The mean μ_i of the ith segment is calculated and subtracted from each sample in that segment.

$$\begin{aligned} \mu_i &= \frac{1}{M}\sum_{n=0}^{M-1} x_i(n) \\ x_i'(n) &= x_i(n) - \mu_i \end{aligned}$$

- The zero-mean segment of data $x_i'(n)$ is then multiplied by a suitable data window $w(n)$, chosen from one of the data windows used in ordinary spectral analysis (e.g. boxcar, Hamming or Hanning). This provides some control over the effects of spectral leakage.

$$x_i''(n) = w(n)x_i'(n) \tag{3.5}$$

- Within each segment, compute the DFT $X_i(k)$;

$$X_i(k) \triangleq \frac{1}{M}\sum_{n=0}^{M-1} x_i''(n)e^{-j2\pi nk/M}, \tag{3.6}$$

where k is discrete frequency. From this DFT the *raw* spectral estimates $\hat{P}_i(k)$ and bispectral estimates $\hat{B}_i(k,l)$ can be formed[7]

$$\begin{aligned} \hat{P}_i(k) &= X_i(k)X_i^*(k), \\ \hat{B}_i(k,l) &= X_i(k)X_i(l)X_i^*(k+l). \end{aligned}$$

- The raw estimates from all K segments can then be averaged to give the following estimates

$$\hat{P}(k) = \frac{1}{K}\sum_{i=0}^{K-1} P_i(k), \tag{3.7}$$

$$\hat{B}(k,l) = \frac{1}{K}\sum_{i=0}^{K-1} B_i(k,l). \tag{3.8}$$

[7]The symbol $\hat{}$ is used to denote estimated quantities.

3.4.4 Properties of these estimates

The properties of the spectral estimate in Equation 3.7 are well documented (see for example [69]), but the properties of the bispectral estimate in Equation 3.8 are rather more interesting. The properties of this estimator for the different signal models M1 and M2 have been considered in different places in the literature, as Table 3.1 illustrates. Here an attempt is made to draw parallels between these two strands of work.

model number	type of signal	Key references
M1	Stochastic linear MA process	[73, 61, 74]
M2	Harmonics in Noise	[75, 76]

Table 3.1: Key references in the theoretical properties of bispectral estimates.

3.4.4.1 Signals conforming to M1

For signals conforming to M1 it has been shown [74] that the smoothed (i.e. averaged) bispectrum estimator is asymptotically *complex normal* [77] (i.e. the estimates of the real part of the bispectrum are normally distributed, and the estimates of the imaginary part of the bispectrum are normally distributed) and asymptotically *independent* (i.e. the estimate at (k,l) is independent of its neighbours at $(k \pm 1, l \pm 1)$).

It has also been demonstrated [74] that the bispectral estimates are asymptotically *unbiased* (that is, that as the data length increases so the expected value of the estimator $\hat{B}(k,l)$ tends to the true bispectrum $B(k,l)$), but, interestingly, that the asymptotic variance has the property: [74, 78];

$$\mathrm{Var}\hat{B}(k,l) \propto P(k)P(l)P(k+l). \tag{3.9}$$

This poses a real problem with bispectral estimates which does not occur with power spectrum estimates, and it has the following consequence; suppose a signal has more *energy* at the normalised frequencies f_1, f_2, f_1+f_2 than it has at g_1, g_2, g_1+g_2. Then the bispectral estimate $\hat{B}(k,l)$ will have a higher variance at the discrete bifrequency corresponding to (f_1, f_2) than it will have at the discrete bifrequency corresponding to (g_1, g_2), purely because of the energy difference, and irrespective of any interesting higher order properties. Thus although a high bispectrum estimate is a *necessary* condition for interesting third-order HOS properties, it is not a *sufficient* one.

Since its variance is energy-dependent, the bispectral *estimate* measures both 2nd order properties as well as 3rd-order properties. This situation is clearly unsatisfactory, and a way around it is to try to remove the second order sensitivity (Equation 3.9) of the estimate. One way in which to do this would be to pre-whiten the signal prior to bispectral analysis [79], but a more common way is to propose the new measure [74]

$$\begin{aligned} s(k,l) &\triangleq \frac{E[B(k,l)]}{\sqrt{E[P(k)]E[P(l)]E[P(k+l)]}} \\ \Rightarrow s^2(k,l) &\triangleq \frac{|E[B(k,l)]|^2}{E[P(k)]E[P(l)]E[P(k+l)]} \end{aligned} \tag{3.10}$$

where $s^2(k,l)$ is called the *skewness function* [74](for reasons which will become apparent later).

If the denominator of Equation 3.10 was known exactly, then the estimate of the skewness function given by

$$\hat{s}^2(k,l) \triangleq \frac{|\hat{B}(k,l)|^2}{P(k)P(l)P(k+l)} \tag{3.11}$$

would have a flat-variance, which follows from Equation 3.9. However, in practice the power spectral terms $P()$s are **not** known, and they too have to be estimated from the data using Equation 3.7. Thus $s^2(k,l)$ in Equation 3.10 is estimated using

$$\boxed{\hat{s}^2(k,l) \triangleq \frac{|\hat{B}(k,l)|^2}{\hat{P}(k)\hat{P}(l)\hat{P}(k+l)}} \tag{3.12}$$

Fortunately, the power spectra estimates $\hat{P}()$ generally have lower variance than the bispectra estimates $\hat{B}()$, and in practice, in the cases where it can be checked, Equation 3.12 often turns out to be very close to Equation 3.11.

A further comment should be added about some confusion which exists in the literature concerning the skewness function. It has been stated [63, 78] that the skewness function is not really a HOS measure because it is a mixture of second- and third-order statistics. This is indeed true from inspection of Equations 3.10-3.12, but it is hoped that the way in which the skewness function has been introduced here makes it clear that the only reason for dividing by the power spectra is to remove the undesirable variance properties of the bispectral estimator. In other words, the normalisation defined in Equation 3.10 makes most sense when written in terms of *estimates* from finite data records (Equation 3.12), since it is in the estimation that the need for normalisation

arises.

3.4.4.2 Signals conforming to M2

The bispectra of signals which conform to a sinusoidal model (M2) do not, in general conform to the statistical properties described above. The mean and variance of the bispectral estimates in this case are very much dependent on the relationships between the frequencies and phases of the component sinusoids. Exact expressions for the bispectra of such signals in the *noise-free* case have been derived which take account for the leakage effects of the data windows [75]. Some of these results will be used in Chapter 4 to derive expressions for the noisy case, but it is relevant to note here that for signals of this type an alternative normalisation is often used [72, 34]

$$b^2(k,l) \triangleq \frac{|E[B(k,l)]|^2}{E[S(k,l)]E[P(k+l)]}, \tag{3.13}$$

in which

$$S(k,l) \triangleq |X(k)X(l)|^2.$$

This can be estimated using

$$\boxed{\hat{b}^2(k,l) \triangleq \frac{|\hat{B}(k,l)|^2}{\hat{S}(k,l)\hat{P}(k+l)},} \tag{3.14}$$

in which

$$\hat{S}(k,l) \triangleq \frac{1}{K}\sum_{i=0}^{K-1} |X_i(k)X_i(l)|^2.$$

This normalised bispectrum, which is called the *squared bicoherence*, does not have the same approximately-flat variance that the skewness function has. However, it does have the useful property that it is bounded between zero and unity, i.e.

$$0 \leq \hat{b}^2(k,l) \leq 1, \tag{3.15}$$

a property which $\hat{s}^2(k,l)$ **does not share**. In fact in practice $\hat{s}^2(k,l)$ and $\hat{b}^2(k,l)$ do usually take very similar values, a fact which has led to some confusion between them in the literature. The next section will present a brief survey of the terminology used for bispectral normalisations, including some other normalisations which have been proposed.

3.5 Bispectrum Normalisation

Table 3.2 summarizes some of the normalisation schemes which have been proposed for the bispectrum. Note that the word "normalisation" is used in a very liberal sense here, since some of the measures, notably $s^2(k,l)$, can exceed unity.

symbol	estimator (Eqn.)	upper bound	signal model	name used here	other names used
s^2	3.10	not known	M1	skewness [74]	
s	3.10	not known	M1	-	3rd order coherency [63], bicoherency [63, 80], bicoherence [68]
b	3.14	1	M2	bicoherence [34]	
b^2	3.14	1	M2	-	bicoherence [72]
b^2_{KB}	3.16	1	M2	KB bicoherence [81]	magnitude squared normalised bispectrum [29]
b_{PO}	3.17	1	none specified	Phase only bicoherence [82]	

Table 3.2: Comparison of the properties of the bicoherence and skewness functions.

In addition to the *skewness* $\hat{s}^2(k,l)$ and the *squared bicoherence* $\hat{b}^2(k,l)$ estimators discussed above, the following normalisations have been proposed the measure:

- [81, 29, 28] proposes

$$\hat{b}^2_{\mathrm{KB}}(k,l) = \frac{|\hat{B}(k,l)|^2}{\left(\sum_{i=0}^{K-1} |X_i(k)X_i(l)X_i^*(k+l)|\right)^2} \tag{3.16}$$

  This measure, which is named after Kravtechenko-Berejnoi, has an advantage over the squared bicoherence $b^2(k,l)$ (from Equation 3.13) since the arguments of its denominator are symmetric under permutations [81]. However as this normalisation has not yet been used extensively, it will not be considered further in this thesis.

- A further normalisation scheme has been suggested [82] in which the bicoherence is decomposed into a *phase-only* measure $b_{\mathrm{PO}}(k,l)$ and an *amplitude-only* measure $b_{AO}(k,l)$. The motivation behind using these measures is that $b_{PO}(k,l)$ contains

only information about the signal phase (i.e. it is not influenced by amplitude fluctuations in the signal of interest). It is estimated using:

$$\hat{b}_{\mathrm{PO}}(k,l) = \frac{1}{K}\sum_{i=0}^{K-1}\frac{X_i(k)X_i(l)X_i^*(k+l)}{|X_i(k)X_i(l)X_i^*(k+l)|}. \tag{3.17}$$

In practice there is often little to choose between these measures, especially between b^2, s^2 and b^2_{KB} because they all have the same numerator, and because of the relative statistical stability of their denominators in comparison with their numerators. Because of this, and in order to keep arguments clear, the main focus in this thesis will be on just one of these normalisation schemes - the squared bicoherence $b^2(k,l)$. Where significant differences between this and the other normalisations are known, they will be mentioned.

3.6 Properties of Normalised Bispectra

In this section some of the useful properties of normalised bispectra will be described. Some of these properties arise from the M1 model, and some from the M2 model, and both models will be of interest in the current study. Many of these properties can be demonstrated by simple proofs, which are presented in the Appendices for completeness. Not all of these properties have been explicitly described in the literature, but those that have are referenced.

Table 3.3 summarizes these properties of the unnormalised and normalised bispectra.

3.6.1 Gaussian Signals

Property 1 *The theoretical bispectrum of a Gaussian signal is identically zero.*

It has been demonstrated (in Appendix A.1) that the cumulants of orders three and higher of a Gaussian process are, in theory, identically zero. Since the bispectrum is related to the third-order cumulant sequence by the Fourier Transform, it follows that the bispectrum of a Gaussian process is identically zero. In fact the bispectrum is zero for any iid signal with a symmetric pdf, not just for Gaussian signals[8]. Following a

[8] It has recently been shown that signals which are not white *can* have nonzero bispectra even if they

contrary argument, signals which have an asymmetric pdf (*skewed* signals) in general have non-zero bispectra [9].

Property 2 *The theoretical bispectrum of a linearly filtered Gaussian signal is identically zero.*

This property follows from the fact that signals formed by linear-filtering Gaussian noise are themselves Gaussian, and so Property 1 applies.

Property 3 *The theoretical bicoherence of a Gaussian signal is zero.*

This follows from Property 1 above, since the denominators of the skewness and bicoherence functions scale the bispectra. It can be shown [63] that these normalised bispectra (the skewness function and the squared bicoherence) have a direct relationship with the signal's third-order moment γ_3 (or "skewness"), and this property has been used by several authors [60, 74] to devise tests for "Gaussianity" [10]. This property, and the statistical tests which are derived from it, are discussed in Appendix A.2.

3.6.2 Additive Noise

Property 4 *The theoretical bispectrum of a non-Gaussian signal is "blind" to additive Gaussian noise [37].*

This follows from Property 1 above. In theory, adding Gaussian noise to a signal will not affect its bispectrum. This theoretical "blindness" to Gaussian noise has been the prime motivation to much of the exploitative HOS research to date.

Property 5 *The theoretical bicoherence of a signal conforming to either the M1 or M2 models (see Section 3.4.1) is, in general,* **not** *"blind" to Gaussian noise.*

---

have symmetrical distributions, and so the bispectra of such signals may still contain useful information [83].

[9] The estimated bispectra of real-life signals are generally non-zero because the estimators used have finite variance unless the data length is infinite. This follows even if the signals have symmetric pdfs.

[10] Note that, subject to the qualifications made elsewhere [83], these tests are in fact for symmetric pdfs, a wider class of signals than Gaussian signals.

This is an important property which has often been neglected by researchers applying bispectral analysis. Put simply, the bicoherence of a signal will be reduced if additive Gaussian noise is added to the signal, because although the bispectrum estimator is theoretically insensitive to Gaussian noise, the denominator terms in Equations 3.12 and 3.14 are not. This is demonstrated in a general way for the skewness function in Appendix A.3, and in a more detailed way for the bicoherence of M2 signals in Appendix A.5.2.

3.6.3 Linear Filters

Property 6 *If a signal is filtered by a linear filter, then, provided that the filter has no zeros on the unit circle, the magnitude of the normalised bispectrum is unchanged.*

Proof : see Appendix A.4.1. This property will be of interest in Chapter 6 when preprocessing of speech signals is investigated.

Property 7 *The theoretical skewness function*[11] *of linearly filtered non-Gaussian iid signals is flat.*

This is demonstrated in [74, 60] and leads to formulations for statistical linearity tests.

Property 8 *If a signal is filtered by a linear phase filter, then its biphase information is unchanged.*

Proof : see Appendix A.4.2. This property will prove to be of interest in Chapter 4 during the consideration of nonlinearity detectors which use the biphase.

3.6.4 Nonlinear filters

Property 9 *The theoretical skewness function*[12] *of a non-Gaussian M1 signal which has been passed through a nonlinear filter may not be flat.*

This property is complementary to Property 7, and is used in the formulation of statistical linearity tests [74, 60].

[11] It is not known whether this result holds for the squared bicoherence function.

[12] It is not known whether this result holds for the squared bicoherence function.

Property 10 *The theoretical bicoherence of an harmonic M2 signal peaks if the signal phases ϕ_1, ϕ_2 and ϕ_3 at frequencies $f_1, f_2, f_3 = f_1 + f_2$ resepectively have the relation $\phi_3 = \phi_1 + \phi_2$.*

This sort of phase relation is known as *Quadratic Phase Coupling* (QPC), and it is an indicator of nonlinear signal generation mechanisms. Under certain estimation conditions phase relations other than $\phi_3 = \phi_1 + \phi_2$ can cause bicoherence peaks, and this problem forms a central part of Chapter 4.

Signal model (input)	Operation	Interesting Properties		Property reference
		unnormalised	normalised	
M1 (Gaussian)	-	$B = 0$	$s^2 = 0$	1, 3
	linear filter			1, 2, 3
M1 (non-Gaussian)	-	$B \neq 0$	$s^2 > 0$	1
	linear filter	-	s^2 flat	7
M1	add Gaussian noise	B invariant	s^2 reduced	4, 5
M2			b^2 reduced	4, 5
M2	linear filter (with no zeros on Unit circle)	B may change	b^2 unchanged	6
M1 or M2	linear phase filter	$\angle B$ invariant	n/a	8
M1 (non-Gaussian)	nonlinear filter	-	s^2 not flat	9
M2	$(\phi_3 = \phi_1 + \phi_2)$	-	b^2 peaks	10
M2	$(\phi_3 \neq \phi_1 + \phi_2)$	-	b^2 does not peak	10

Table 3.3: Theoretical properties of un-normalised and normalised bispectral estimators. For clarity the indices (k, l) have been dropped from the bispectral terms.

3.7 Estimation Issues

The main estimation parameters which need to be chosen for bispectral analysis are the same parameters required for spectral analysis. However, whereas for spectral analysis the interpretation of effects due to windowing, noise and insufficient data length are well understood, the situation with normalised bispectra is quite different. In many

ways the normalised bispectra have more in common with the second order coherence function than with the power spectrum [84], and as such some problems which occur in coherence analysis can occur here too. In this section some of these matters are addressed, and once again an attempt is made to bring together the disparate literature concerning M1 and M2 models.

3.7.1 Choice of Data Window

The issue of data windowing for normalised bispectrum analysis has not yet been addressed in the literature. For M1 signals no analysis has ever been made. For M2 signals a major development in this direction was reported in [75], where an expression for the (unnormalised) bispectrum of an harmonic signal (in zero noise) was derived. Using this result, together with the expression for the DFT of the harmonic signal, an expression can be determined for the normalised bispectrum of the signal. The special case of Equation 3.4 is considered in which $P = 3, f_3 = f_1 + f_2, \phi_3 = \phi_1 + \phi_2, v(n) = 0$, as this will be of special interest in Chapter 4.

Theorem 1 *The theoretical normalised bispectrum* [13] *of an M2 signal (Equation 3.4) with sinusoids of unit amplitude and zero noise, estimated using a data window with Dirichlet kernel $D(k, f_p, M)$, is given by Equation 3.10 :*

$$s^2(k,l) \triangleq \frac{|E[B(k,l)]|^2}{E[P(k)]E[P(l)]E[P(k+l)]}$$

in which

$$\begin{aligned} E[\hat{B}(k,l)] &= \sum_{p,q,r,=1}^{3} \frac{1}{8} \left[\sum_{a,b,c=0}^{1} D(k,(-1)^a f_p, M) D(l,(-1)^b f_q, M) D(k+l,(-1)^{c+1} f_r, M) \times \right. \\ &\quad \left. e^{-j\pi(M-1)((-1)^a f_p + (-1)^b f_q + (-1)^c f_r)} E\left[e^{j((-1)^a \phi_p + (-1)^b \phi_q + (-1)^c \phi_r)}\right] \right], \\ E[\hat{P}(k)] &= \sum_{p=1}^{3} \frac{1}{4} \left[D^2(k, f_p, M) + D^2(k, -f_p, M) \right] \end{aligned}$$

where $D(k, f_p, M)$ is the Dirichlet kernel of the window, which for the boxcar window is given by

$$D(k, f_p, M) = \frac{\sin\{\pi(k - f_p M)\}}{M \sin\{\pi(k - f_p M)/M\}}.$$

[13] Both skewness s^2 and squared bicoherence b^2 reduce to the same expression, but the skewness form is given here as it is slightly simpler.

The kernels associated with other windows are described in [75].

Proof : see Appendix A.5.1.

This result is compared with empirical results (from simulations, with $N = 65536$ and $K = 1024$) in Figures 3.6 and 3.7 for the DFT size $M = 64$. Figure 3.6 shows the excellent agreement between the theoretical (from Equation A.11) and empirical power spectra, except at very low frequencies. This difference is due to the fact that during the spectral estimation the mean is subtracted from each frame. It is evident that of the three windows shown the boxcar (or rectangular) window has the narrowest mainlobe, but the highest sidelobe levels.

Figure 3.7 shows the theoretical and empirical squared bicoherences of the same signal. It is clear that again the boxcar window results in the highest sidelobe levels.

It is interesting to note the large difference between the squared bicoherence estimated using a Hamming window and that estimated using a Hanning window (in Figure 3.7). In (power) spectral analysis both windows give similar results, although the sidelobe structure for the Hamming window is less smooth than the Hanning window (Figure 3.6). These features become more pronounced in the squared bicoherence (Figure 3.7), and the Hanning window estimate has a very wide main lobe, but a smooth shape, whereas the Hamming window estimate has a narrower main lobe but a less smooth shape.

In order to quantify how well each of these windows performs at resolving the peak corresponding to the normalised bifrequency (f_1, f_2), the following measure is proposed

$$\beta = \frac{\hat{b}^2(f_1, f_2)}{\sum_{IT} \hat{b}^2(k, l)} \times 100\%.$$

β is thus the squared bicoherence level at the discrete bifrequency which corresponds most closely to the signal sinsuoid frequencies expressed as a percentage of the total squared bicoherence "energy" in the IT. Windows which are good at resolving the peak will have a larger value of β than windows which are poor at resolving the peak.

Table 3.4 compares the values of β for the three windows shown in Figures 3.6 and 3.7, for the theoretical and the empirical squared bicoherences (β_{theory} and β_{emp} respectively). It is evident that although there are some differences between the empirical results and the theoretical results, the Hamming window is most successful at resolving

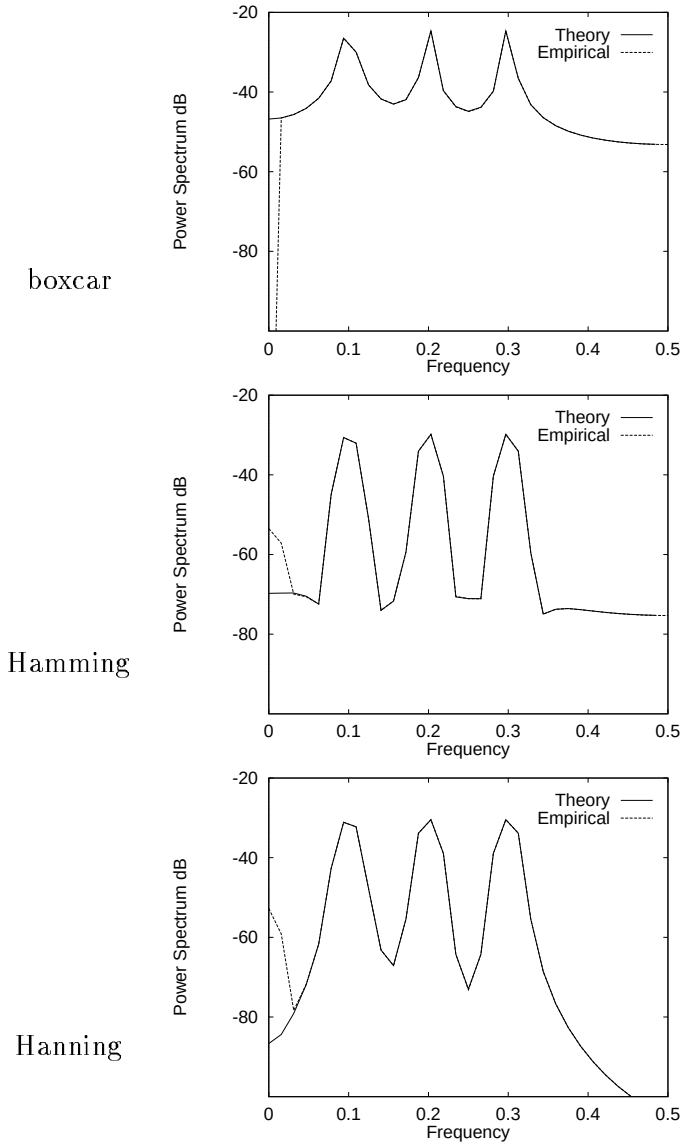

Figure 3.6: Theoretical and empirical power spectra estimates for signal composed of three sinusoids in noise with different data windows: boxcar (top), Hamming (middle) and Hanning (bottom).

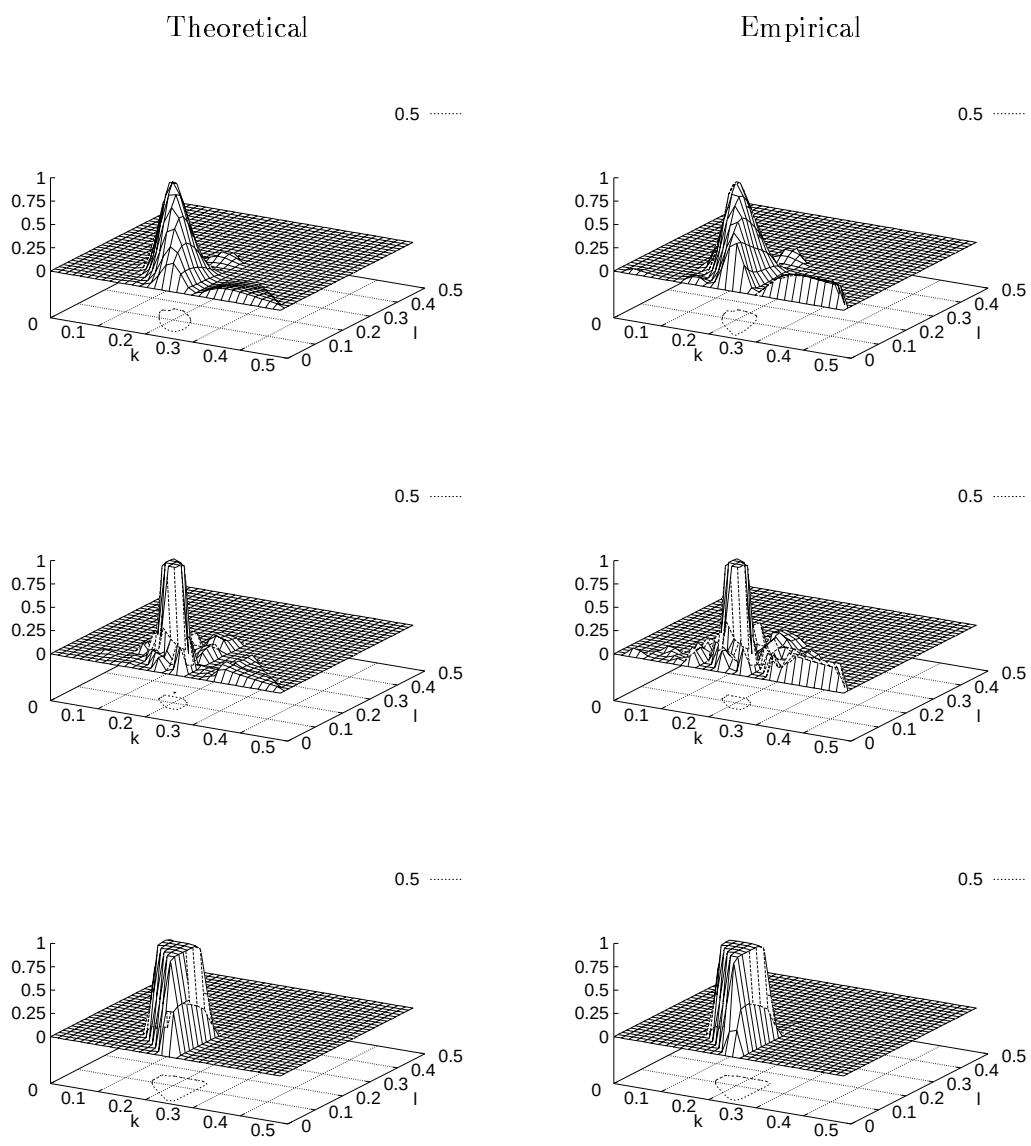

Figure 3.7: Theoretical (left) and empirical (right) squared bicoherence for signal composed of three sinusoids in noise with different data windows: boxcar (top), Hamming (middle) and Hanning (bottom). The contour shown is at the 0.5 level.

the peak corresponding to (f_1, f_2).

Window	β_{theory} (%)	β_{emp} (%)
boxcar	2.9	2.4
Hamming	4.2	3.1
Hanning	1.8	2.5

Table 3.4: Comparison of performance measure β for three data windows.

3.7.2 Data length required for good estimates

It is widely known that, for a given data length, bispectral estimates generally have higher variances than power spectral estimates. Therefore the data lengths which are sufficient for reliable power spectrum estimates **may not** be sufficient for good bispectral estimates. A number of researchers have published guidelines and empirical results for reliable estimation of normalised bispectra;

- Hinich et al [85] suggests that, for an M1 signal, if no frequency smoothing is used, the number of segments of data should be at least as large as the DFT size, i.e. $K \geq M$. Collis [66, 86] further extends this idea to trispectral analysis in which $K \geq M^2$.
- Elgar et al [76, 72] produce empirical results for the mean and variance of bicoherence estimates. These are given in terms of b^2, the "true bicoherence[14]".

The problem with these rather general approaches is that neither takes account of the facts which underpin the choice of data length in practical spectral analysis, that is :

The length of data required depends on how noisy the data is.

Elgar's expressions for the mean and variance of the bicoherence estimates (which are given in terms of the *true* bicoherence b^2) [72, 87] are of limited value unless the expected b^2 value is known. This in turn depends (in some as yet undetermined way)

[14]The "true bicoherence" is defined as the proportion of energy at the bifrequency of interest which is due to the determinstic (i.e. sinusoidal) coupled components. In other words the true bicoherence is a measure of how much coupled energy there is at the bifrequency of interest, and it is related to the amplitudes of the sinusoids.

on the SNR. Investigation of this matter leads to the following Theorem[88], which is a simplified special case of a more general result (see Appendix A.5.2):

Theorem 2 *If the effects of leakage are ignored, then the peak bicoherence $b^2(k,l)$, calculated at the bifrequency (k,l) which corresponds to the coupled components at frequencies f_1, f_2 and f_1+f_2, of an $M2$ signal consiting of three equal-amplitude coupled harmonics in variable levels of additive white Gaussian noise with a signal-to-noise ratio of SNR is given by*

$$b^2(f_1,f_2) = \frac{1}{1+\frac{18}{M}10^{-\mathrm{SNR}/10}+\frac{72}{M^2}10^{-2\mathrm{SNR}/10}}. \tag{3.18}$$

Proof : see Appendix A.5.2.

This result indicates how the theoretical squared bicoherence peak height varies as the SNR and the DFT size vary. In practice the actual value of the SNR is usually beyond the experimenter's control, although it can be estimated from the power spectrum. The DFT size M is a parameter chosen by the experimenter.

Figure 3.8 shows the evaluation of Equation 3.18 over a range of SNRs, and with a variety of DFT sizes M. The dependence of the peak b^2 value on M arises because, as the DFT size increases, so the harmonics appear more clearly in the frequency bins of interest in the DFT. For a given M the squared bicoherence peak will be close to 1 for high SNRs but if the SNR is reduced then at a certain threshold level b^2 will fall rapidly.

Now Equation 3.18 becomes really useful when the empirical expressions for the bias and variance of the bicoherence estimate (from [72]) are considered. These expressions are [72]

$$\begin{aligned} \mathrm{Bias}\,\hat{b}^2 &\approx \left(\frac{1}{K}\right)\left(1-b^2(k,l)\right)^2, \\ \mathrm{Var}\,\hat{b}^2(k,l) &\approx \left(\frac{2b^2(k,l)}{K}\right)[1-b^2(k,l)]^3. \end{aligned} \tag{3.19}$$

Now replacing $b^2(k,l)$ in these expressions with the expression in Equation 3.18 gives relations between the bias and variance of the bicoherence estimate in terms of the SNR, the DFT size and the number of data segments K, and these can be used to give some idea of the choices of parameters available for analysis of signals with known

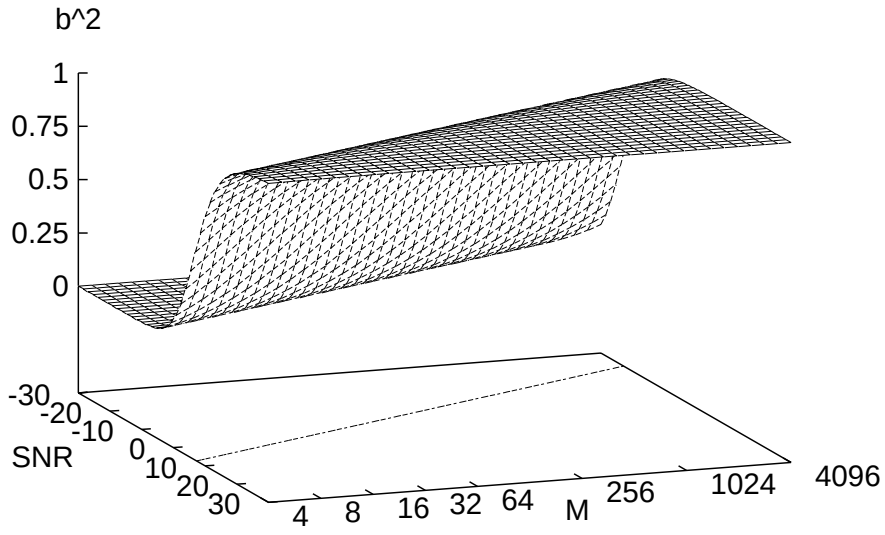

Figure 3.8: Theoretical results showing how squared bicoherence level, at the single discrete bifrequency which corresponds to the signal's coupled sinusoids, varies with SNR for different DFT sizes M. The contour is at the $b^2 = 0.5$ level.

SNRs. It is desirable to have the DFT size as large as possible, since this gives more frequency resolution, but this requirement conflicts with the requirement of having a large number of segments K to give reliable estimates.

The variation of estimator bias and variance for different segment sizes and different SNRs, are shown in Figures 3.9 and 3.10 respectively. As expected the bias and variance decrease as K increases, a result which is consistent with the theoretical result that the estimator is asymptotically *unbiased* and *consistent*). The bias increases as the noise level increases, as expected, but it can be seen that as long as K is not too small, the bias is very low over most SNRs. The variance varies in a more complicated way, but again it can be noted that, as long as K is not very small, that the variance is quite low.

The fact that the variance of the estimator seems to be very low at low SNRs (the part labelled "A" in Figure 3.10) is thought to be due to the following; at very low SNRs the bicoherence level is very small, so even though the variance appears to be very low at low SNRs, it is in fact quite large in comparison with the bicoherence level.

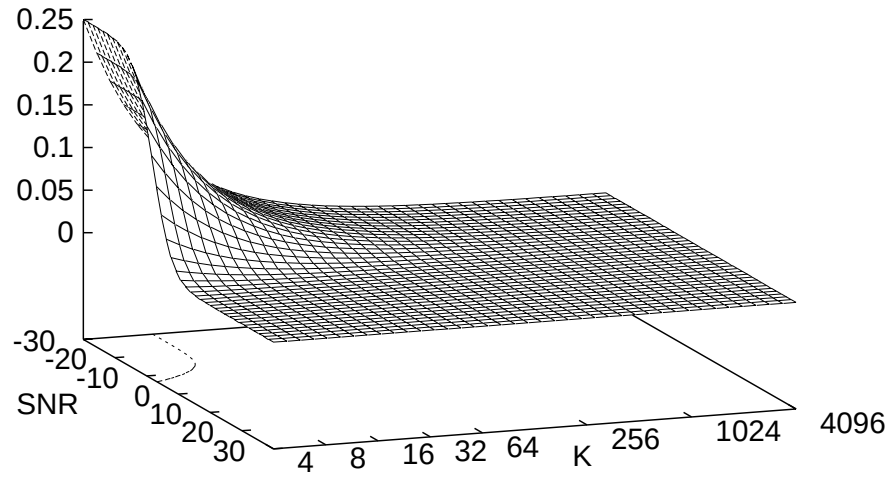

Figure 3.9: Theoretical results showing how bias of squared bicoherence estimator (vertical axis) varies with number of data segments K and with SNR. The contour is at the 0.1 level.

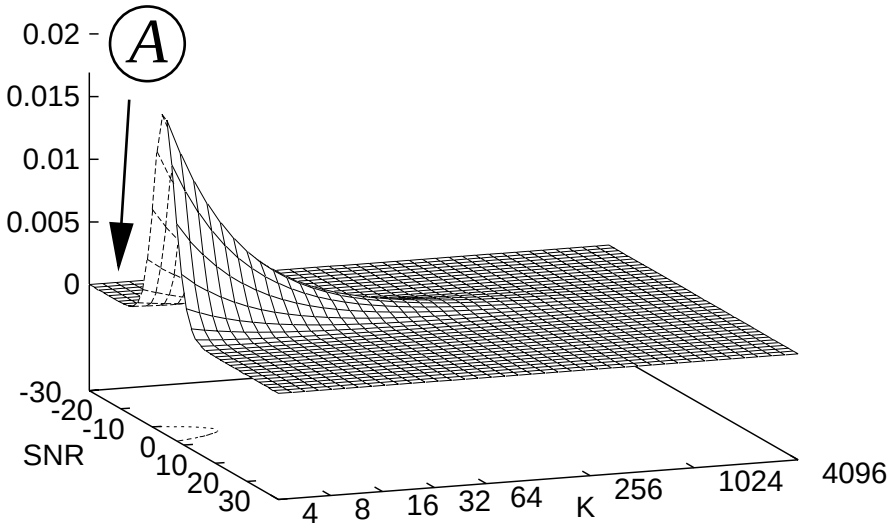

Figure 3.10: Theoretical results showing how variance of squared bicoherence estimator (vertical axis) varies with number of data segments K and with SNR. The contour is at the 0.01 level.

3.8 Summary

This chapter began by taking a statistical viewpoint of signal characterisation, considering the description of a stochastic signal by its moments or cumulants. The bispectrum was introduced as a frequency domain representation of the third-order cumulant, and the statistical assumptions on which this result were developed were then relaxed to establish an engineering framework for bispectrum estimation. Bispectral normalisations were introduced as solutions to estimation problems, and a variety of normalisations were described. The properties of these measures were then described in some detail, with particular attention paid to the bicoherence of harmonic signals. New expressions were derived for the bicoherence of such signals, and the effects of data windows were investigated. Finally, a consideration of estimator properties in noise led to some useful expressions which give an indication of how the estimate will perform on a given data length, at a given SNR.

Chapter 4

Quadratic Phase Coupling Detection

4.1 Introduction

In Chapter 3 a harmonic signal model M2 was introduced (Equation 3.4), and some of the properties of its bicoherence[1] were investigated. The reason why this model holds interest is because it is closely related to a simple nonlinear signal generation mechanism, which will now be described.

Figure 4.1 illustrates a simple filter composed of linear and nonlinear components. If the input to this system $y(n)$, is a sum of two sinusoids (with normalised frequencies f_p);

$$y(n) = \sum_{p=1}^{2} \cos(2\pi f_p n + \phi_p), \tag{4.1}$$

then the output will be given by

$$\begin{aligned} x(n) &= y(n) + y^2(n) && (4.2)\\ &= \cos(2\pi f_1 n + \phi_1) + \cos(2\pi f_2 n + \phi_2) + \\ &\quad [\cos(2\pi f_1 n + \phi_1) + \cos(2\pi f_2 n + \phi_2)]^2 \\ &= \cos(2\pi f_1 n + \phi_1) + \cos(2\pi f_2 n + \phi_2) + \cos^2(2\pi f_1 n + \phi_1) + \\ &\quad 2\cos(2\pi f_1 n + \phi_1)\cos(2\pi f_2 n + \phi_2) + \cos^2(2\pi f_2 n + \phi_2). && (4.3) \end{aligned}$$

[1] Following Table 3.2 the bicoherence is the measure used in analysis of M2 models, although much of what is discussed in this Chapter applies also to the skewness function.

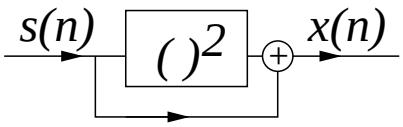

Figure 4.1: Schematic diagram of simple generator for QPC signal.

Now applying the trigonometric identities

$$\begin{aligned} \cos^2 A &= \frac{\cos 2A + 1}{2} \\ \cos A \cos B &= \frac{1}{2}[\cos(A+B) + \cos(A-B)] \end{aligned}$$

to Equation 4.3 gives

$$\begin{aligned} x(n) &= \cos(2\pi f_1 n + \phi_1) + \cos(2\pi f_2 n + \phi_2) + \\ &\quad 1 + \frac{1}{2}\cos(4\pi f_1 n + 2\phi_1) + \cos(2\pi(f_1+f_2)n + (\phi_1+\phi_2)) + \\ &\quad \cos(2\pi(f_1-f_2)n + (\phi_1-\phi_2)) + \frac{1}{2}\cos(4\pi f_2 n + 2\phi_2). \end{aligned} \tag{4.4}$$

Clearly the output signal $x(n)$ has a richer harmonic structure than the input signal, since it now contains components with frequencies 0 (DC), $2f_1$, $2f_2$, f_1+f_2 and f_1-f_2. This harmonic richness is a property of signals produced by passing harmonic signals through nonlinear filters [3].

Equation 4.4 also indicates that as well as having a particular harmonic structure, the components of $x(n)$ also have related *phases*, as shown in Table 4.1. Since these relationships between the phases arise from a quadratic nonlinearity in the signal production mechanism, the signal is said to exhibit "Quadratic Phase Coupling" (QPC)[34]. Now

frequency	phase
$2f_1$	$2\phi_1$
$2f_2$	$2\phi_2$
f_1+f_2	$\phi_1+\phi_2$
f_1-f_2	$\phi_1-\phi_2$

Table 4.1: Phase relations between frequencies for nonlinear signal.

referring back to Chapter 3, the definition of the bispectrum (from Equation 3.2)

$$B(k,l) = X(k)X(l)X^*(k+l),$$

can be rewritten in magnitude and phase form ($X(k) = |X(k)|e^{j\phi(k)}$) as

$$\begin{aligned} |B(k,l)| &= |X(k)X(l)X^*(k+l)| \\ \angle B(k,l) &= \phi(k) + \phi(l) - \phi(k+l). \end{aligned}$$

It becomes apparent how the bispectrum might be useful in testing for this type of nonlinearity, since the following relations occur;

$$\begin{aligned} \angle B(f_1, f_1) &= \phi_1 + \phi_1 - 2\phi_1 = 0 \\ \angle B(f_2, f_2) &= \phi_2 + \phi_2 - 2\phi_2 = 0 \\ \angle B(f_1, f_2) &= \phi_1 + \phi_2 - (\phi_1 + \phi_2) = 0 \\ \angle B(f_1, -f_2) &= \phi_1 - \phi_2 - (\phi_1 - \phi_2) = 0, \end{aligned} \tag{4.5}$$

where the notation $B(f_1, f_2)$ refers to the discrete bifrequency (k, l) which corresponds most closely to the normalised bifrequency (f_1, f_2).

Thus signals which are generated by nonlinear models of the type shown in Figure 4.1 will have zero biphase at bifrequencies related to the frequencies present in the input. This raises the possibility of testing a signal's bispectrum for zero biphase to see whether the signal's generation could have involved a quadratic nonlinearity.

In the next section a useful interpretation of this property in the complex bispectral plane will be given, as this then allows development of QPC detectors based on the bicoherence. But first it will be helpful to set out QPC detection in a hypothesis testing framework, by defining two signals with *the same* power spectra, but *different* QPC properties. For convenience, these signals will be generated by the M2 model, which is a simple sum of sinusoids, and so the quadratic nonlinearity is only *implicit*. The form of this model is easier to manipulate than the explicitly nonlinear model (shown in Figure 4.1), and by careful choice of the model parameters it can easily generate the two types of signals of interest; those which *do* exhibit the signatures of nonlinearity (frequency- and phase-coupling); and usefully, those which *do not*.

4.2 Two test signals

Two types of M2 signal (from Equation 3.4) will be considered, with $P = 3$:

Uncoupled : M2(UC) $\phi_3 \neq \phi_1 + \phi_2$

Coupled : M2(QPC) $\phi_3 = \phi_1 + \phi_2$

Both signals have identical power spectra, with peaks at corresponding to the signal energy at f_1, f_2 and $f_1 + f_2$, but whereas the M2(QPC) signal exhibits QPC, the M2(UC) signal does not. It is stressed here that using only second order techniques it is impossible to distinguish between M2(UC) and M2(QPC) signals.

So the problem is : presented with a signal with energy at the three frequencies f_1, f_2 and $f_1 + f_2$, can the appropriate signal model, M2(QPC) or M2(UC), be determined ? This problem can be rephrased as a choice between two hypotheses : the hypothesis that the unkown signal exhibits QPC, and the hypothesis that the signal does not exhibit QPC.

In the remainder of this chapter, techniques for choosing between these hypotheses are considered. For now it will be assumed that when estimating the bispectrum, the phases of the different segments are *independent* of each other, so the uncoupled case is in fact characterised by $\phi_3 = U[0, 2\pi)$. The effect of breaching this assumption will be investigated in Section 4.4. The next section introduces a new way of looking at the bispectrum which is helpful in understanding how the bispectrum can be used to detect QPC.

4.3 Phasor Interpretation of QPC

The bispectrum $B(k,l)$ is a complex quantity, having magnitude and phase. Consequently for each (k,l) the bispectrum value can be represented as a point in complex $\Re[B(k,l)]v.\Im[B(k,l)]$ space. Figure 4.2 shows the bispectrum phasor at the discrete bifrequency corresponding to (f_1, f_2) for the M2(UC) and the M2(QPC) models described above. The length of the phasor is the bispectrum magnitude at the bifrequency corresponding to (f_1, f_2), and the angle the phasor makes with the positive real axis is the biphase. It is evident that in the case of phase coupling the biphase is zero.

Using this interpretation, the QPC-detecting properties of the bispectrum (and bicoherence) can be explained. Consider two signals with the properties of M2(UC) and M2(QPC) described above. Assume that the bispectrum is computed by averaging the segmental raw bispectrum estimates over these K segments. As there will be K raw bispectra computed there will be K biphase phasors at each bifrequency, one for each

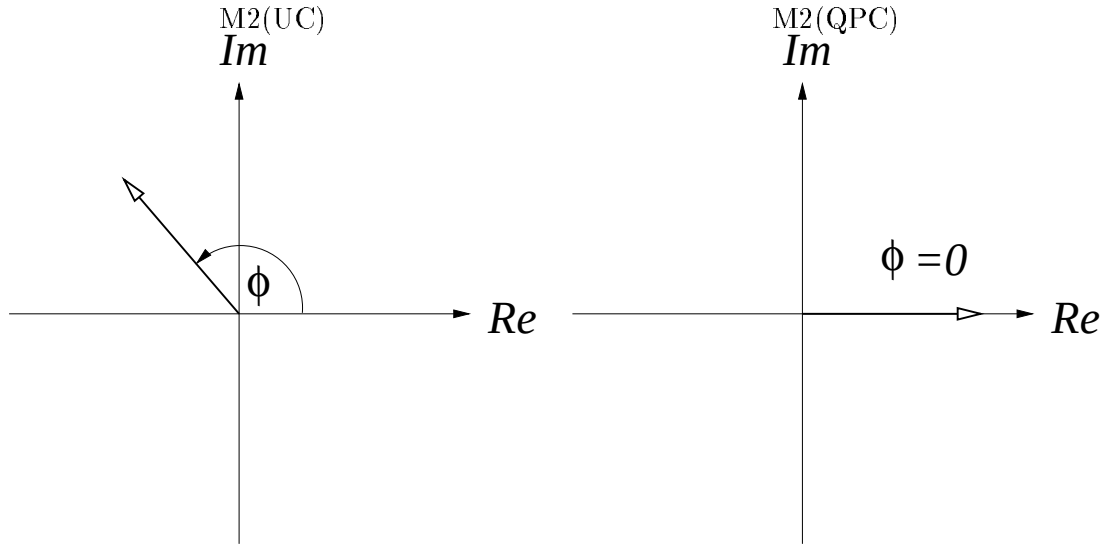

Figure 4.2: Schematic diagram of phasor representation of bispectra; Uncoupled case M2(UC) (left) and coupled case M2(QPC) (right).

data segment. Now focus attention on these phasors at the discrete bifrequency which corresponds to the sinusoid frequencies, which is shown schematically in Figure 4.3.

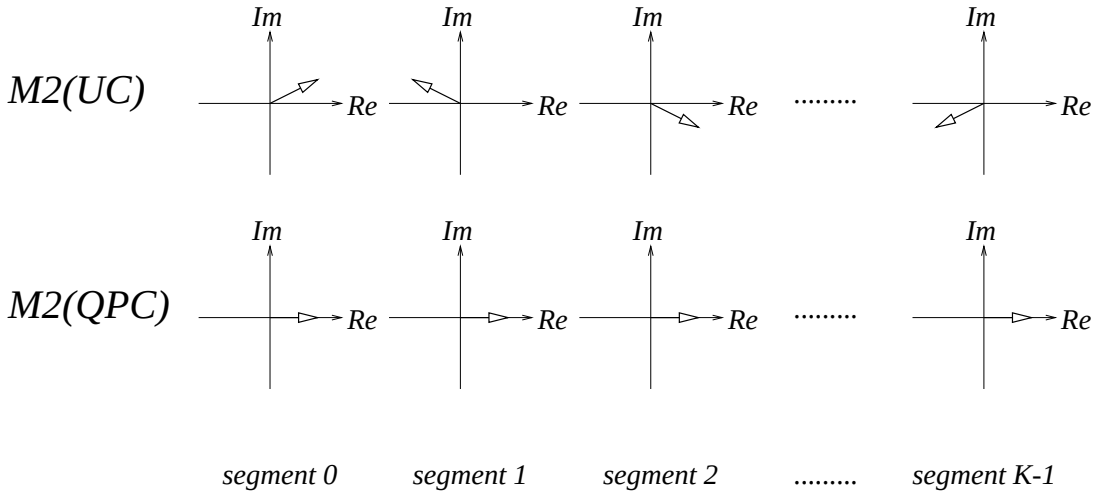

Figure 4.3: Schematic representation of biphase phasors at bifrequency corresponding to (f_1, f_2) for Uncoupled [M2(UC)] and Coupled [M2(QPC)] signals for K data segments.

- For the M2(UC) signal, each of the K phasors will be oriented randomly over 2π. The phasors will sum *incoherently*, to give an average which tends to zero as K increases.
- For the M2(QPC) signal, each of the K phasors will be oriented along the positive real axis. The phasors will sum *coherently*, to give a non-zero average.

Thus the bispectrum estimate for the M2(UC) signal will be zero, but for the M2(QPC) signal will be nonzero.

The bicoherence is simply a normalised bispectrum measure, and so the mechanism behind its use as a QPC detector is identical to the one presented here for the bispectrum. The magnitude of the squared bicoherence $b^2(k,l)$ can be interpreted[34] as the *proportion* of energy at the frequency $k+l$ which is coupled with the components at frequencies k and l. Thus if b^2 is 1 then there is total coupling and if b^2 is 0 there is no coupling.

This interpretation can be understood in terms of Equation A.19 from Appendix A.5.2, reproduced here;

$$\begin{aligned} b^2(f_1,f_2) &= \frac{\frac{A^6}{64}}{\left[\frac{A^4}{16}+\frac{\sigma_v^2}{M}\frac{2A^2}{4}\right]\left[\frac{A^2}{4}+\frac{\sigma_v^2}{M}\right]} \\ &= \frac{1}{1+\frac{12}{M}\frac{\sigma_v^2}{A^2}+\frac{32}{M^2}\frac{\sigma_v^4}{A^4}} \end{aligned} \tag{4.6}$$

This equation indicates how the squared bicoherence at the discrete bifrequency corresponding to the sinusoid frequencies f_1, f_2 depends on the amplitudes of harmonics A, which determine the amount of signal energy which is coupled[2], and the background noise variance σ_v^2, which determines the amount of signal energy which is uncoupled - it is evident that if the proportion of uncoupled energy increases (because σ_v^2 increases), so b^2 will decrease. This interpretation will be used in the analysis of the speech database in Chapter 7.

4.4 Phase Randomisation

It was assumed in Section 4.2 that the sinusoid phases were random variables over $[0, 2\pi)$. In practical terms, this means that when generating K segments of M2 signal data, the component phases have to be randomised at the start of each segment. This is a matter which has not been widely acknowledged, but is of great practical importance. The mechanism, described above, which leads to the bispectrum's property of being able to distinguish between UC and QPC signals, relies heavily on this assumption, and if this assumption is breached new problems can arise.

[2]It is assumed here that the sinusoids all have the same amplitude A.

Consider an extension of the models considered above. For both M2(UC) and M2(QPC) models, now consider the two cases of i) phase randomisation (PR) and ii) constant phase (CP). This results in four models, whose generation is shown schematically in Figure 4.4, and whose properties are summarised in Table 4.2. The first three columns of the table ("Name", "QPC", "ϕ randomisation ?") describe the properties of the model, and the last three columns ("$|B(f_1, f_2)|$", "$b^2(f_1, f_2)$" and "$\angle B(f_1, f_2)$")[3] are the bispectral properties of such signals, as observed from simulations, and which are explained from a theoretical viewpoint below.

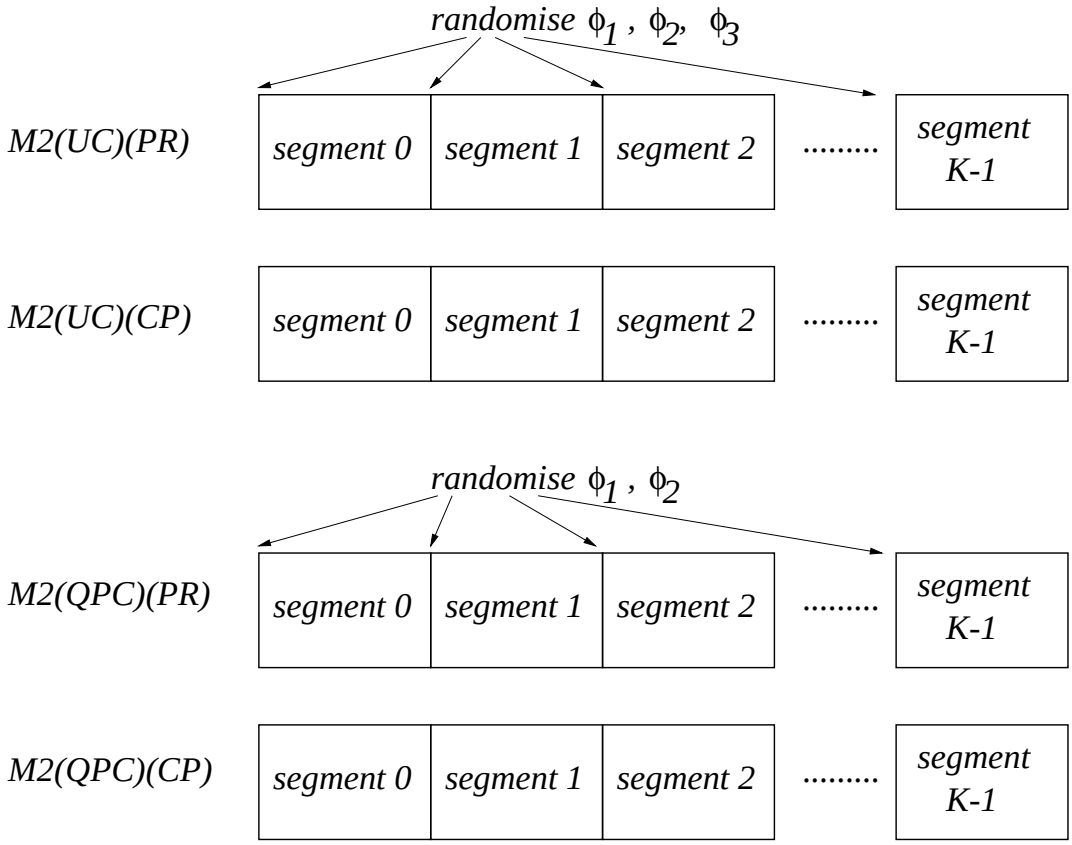

Figure 4.4: Schematic diagram showing generation of four types of M2 signals.

Name	QPC ?	ϕ randomisation ?	$\lvert B(f_1, f_2)\rvert$	$b^2(f_1, f_2)$	$\angle B(f_1, f_2)$
M2(UC)(PR)	no	yes	0	0	$U[0, 2\pi)$
M2(UC)(CP)	no	no	$\neq 0$	$\neq 0$	constant
M2(QPC)(PR)	yes	yes	$\neq 0$	$\neq 0$	0
M2(QPC)(CP)	yes	no	$\neq 0$	$\neq 0$	0

Table 4.2: Properties of four types of M2 signals

Now at (f_1, f_2) the K bispectrum phasors will add incoherently *only* for the M2(UC)(PR) model. The other three models, namely M2(QPC)(PR), M2(QPC)(CP) and M2(UC)(CP)

[3] where (f_1, f_2) is here taken to mean the discrete bifrequency which corresponds to the normalised bifrequency (f_1, f_2).

will have phasors which point in the same direction for each segment, and so add coherently to give a bispectral peak. It is the last of these models, M2(UC)(CP), which causes a problem, since it exhibits no coupling, yet results in a bispectral peak. This leads to the following property of segment averaged bispectral estimates;

Property 11 *If the phase randomisation assumption does not hold, then the bispectrum (and bicoherence) can peak even if there is no QPC.*

Perhaps the easiest way to estimate the bicoherence is by segment averaging, and this technique has been widely used. What the result above indicates is, that if the segmented data frames have the same properties as the M2(UC)(CP) signal, then the bicoherence will be a poor measure of QPC. Consider a simple signal composed of a single sinusoid with frequency f (normalised) and phase ϕ.

It is easy to imagine a signal, perhaps a speech signal, or a machine noise signal, in which the spectrum contains several frequency components whose frequencies are related such that $f_3 = f_1 + f_2$. This is frequency coupling, and it has already been discussed that it is very weak evidence of QPC. However, the bicoherence magnitude of such a signal, estimated using segment averaging, will have a peak corresponding to (f_1, f_2), simply because the biphase of each segment (given by $\phi_1 + \phi_2 - \phi_3$) is *the same* for each segment. This is the danger in interpreting the segment averaged bicoherence magnitude as an indicator of QPC for deterministic signals.

4.4.1 Misinterpretation of bicoherence

From a practical point of view, the literature contains accounts of applications of bispectrum/bicoherence analysis to try to detect QPC in signals where the phase randomisation assumption may not hold. In these accounts, the bicoherence is usually estimated using a segment-averaging approach to a single record of data to a quasi-periodic signal, and significant levels of squared bicoherence are taken as evidence of nonlinearity. As it has been shown, for these signals, estimated in this way, the interpretation of the bicoherence magnitude as a measure of the extent of QPC is erroneous. Examples in which this misinterpretation has occurred include the application of bicoherence[4] to voiced speech[36, 44, 45] and musical signals [46] as well as the author's own work on machine noise [89, 90].

[4] or the skewness function, for which identical arguments apply.

4.5 A New QPC Detector

To get around the problems described above, a new QPC detector will be described which does not rely on the PR assumptions made above. Central to the operation of this detector are the properties shown in the last column of Table 4.2. The bicoherence *magnitude* is unable to distinguish between zero biphase (which indicates QPC) and constant biphase (which does not indicate QPC). The logical improvement is to devise a test for zero biphase, which can be unambiguosly interpreted as a sign of QPC.

The biphase estimate $\hat{\theta}(k,l)$ is calculated from

$$\hat{\theta}(k,l) = \arctan \frac{\Im[\hat{B}(k,l)]}{\Re[\hat{B}(k,l)]}$$

where $\hat{B}(k,l)$ is the simple segment-averaged bispectral estimate. The properties of this estimate have been investigated empirically [72, 87], and it was found that the estimate is approximately normally distributed around the true biphase $N(\theta, \sigma^2)$ with variance[5]

$$\sigma^2[\hat{\theta}(k,l)] \approx \frac{1}{2K}\left(\frac{1}{b^2_{\text{true}}(k,l)} - 1\right) \tag{4.7}$$

where $b^2_{\text{true}}(k,l)$ is the true squared bicoherence, which is to say that it represents the proportion of coupled energy at (k,l)[72]. This is an approximate expression, and it is clear that as the true squared bicoherence approaches zero, so the variance of the biphase estimate, according to Equation 4.7, approaches ∞. In practice the biphase is always bounded between 0 and 2π, so it seems that the approximation breaks down for low levels of true bicoherence. More will be said on this matter in Section 4.5.4 below.

Of course in practice b^2_{true} is not known, so it has to be replaced with $\hat{b}^2_{\text{true}}$, which is estimated from the data. Now, dividing the biphase estimate by its (approximate) standard deviation (from Equation 4.7) gives a normal variable

$$Z(k,l) \approx \frac{\hat{\theta}(k,l)}{\hat{\sigma}[\theta(k,l)]} \approx N(\frac{\theta}{\hat{\sigma}[\theta(k,l)]}, 1), \tag{4.8}$$

and if the true biphase, θ, is zero, then $Z(k,l)$ becomes a standardised normal variable $Z(k,l) \approx N(0,1)$.

[5] It is interesting to note that this relation is similar in form to the expression for the variance of the phase of the second-order coherence function [84].

It follows that a critical level for the phase determined from Equation 4.8 can be used to define a critical region in the complex plane. A hypothesis test can thus be formulated with the following hypotheses

H_0 The biphase is zero (for M2(QPC) signals).

H_1 The biphase is non-zero (for M2(UC) signals).

4.5.1 The complex bicoherence

For convenience a new quantity called the *complex bicoherence* will be introduced, which will make the development of the QPC detector simpler. It is defined as

$$\boxed{b_c(k,l) \triangleq \frac{E[X(k)X(l)X^*(k+l)]}{\sqrt{E[|X(k)X(l)|^2]E[|X(k+l)|^2]}}}$$

and has the following properties;

Phase :

$$\angle b_c(k,l) = \arctan\frac{\Im[b_c(k,l)]}{\Re[b_c(k,l)]} = \angle B(k,l)$$

i.e. it has the same phase properties as the bispectrum.

Magnitude :

$$|b_c(k,l)|^2 = \Re[b_c(k,l)]^2 + \Im[b_c(k,l)]^2 = b^2(k,l), \tag{4.9}$$

i.e. its magnitude squared has the same properties as the squared bicoherence.

The QPC detection problem can now be represented in the complex $b_c(k,l)$ plane rather than the complex $B(k,l)$ plane. This has the advantage that bifrequency bins corresponding to squared bicoherence peaks will now be close to the unit circle ($|b_c(k,l)| \approx 1$).

Testing for zero biphase alone is unsatisfactory, firstly from a mathematical viewpoint, because it can be seen from Equation 4.7 that if the true bicoherence is small the biphase variance is very large, and secondly from a practical viewpoint, because bifrequencies at which the bicoherence is very small are likely to be noise dominated, and so unlikely to hold any information of interest.

To get around this problem, a hybrid two-stage test will now be proposed in which firstly a test for zero bicoherence is carried out to screen out bifrequencies which just contain

noise, and secondly the biphase test is carried out on the remaining bifrequencies.

4.5.2 A two-part test for QPC

Figure 4.5 shows schematically how the QPC works; first the bicoherence (and complex bicoherence) are estimated from the data. Each bifrequency is then tested to see if its magnitude is statistically significant[6], and those bifrequencies which have significant bicoherence magnitude have their phases tested.

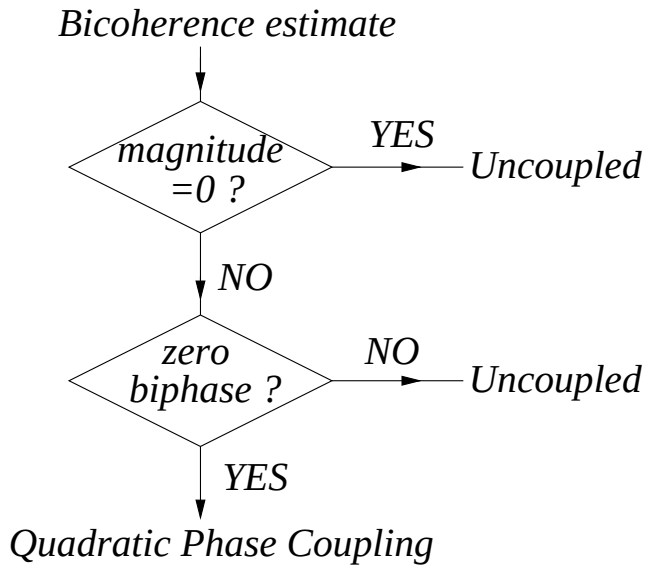

Figure 4.5: Schematic diagram of two-part QPC detector.

4.5.2.1 Testing for significant bicoherence magnitude

The test for significant bicoherence magnitude is intended to remove from future consideration bifrequencies at which there is no signal of interest, and just noise. This is in fact closely related to Hinich's "Gaussianity" test[74][7], but is carried out *at each bifrequency*. The hypotheses tested are:

H_0 The bicoherence at this bifrequency is zero (the bicoherence contains only Gaussian noise at this bifrequency).

H_1 The bicoherence at this bifrequency is non-zero (there might be QPC at this bifrequency).

[6] What is meant by "significant" will be discussed below.

[7] applied to the bicoherence rather than the skewness.

Under H_0, Hinich[74] showed that the scaled *skewness* function $2Ks^2(k,l)$ is approximately asymptotically central-χ^2 distributed with 2 degrees of freedom. Empirical agreement has been reported for the *bicoherence* function [76], and confidence limits for the scaled skewness function can be easily converted to get confidence levels for the complex bicoherence, and this is described in detail in Appendix A.2.

Figure 4.6 shows the contour of equal probability in the complex bicoherence plane, for $\alpha = 0.05$. If the complex bicoherence falls within the shaded circle, then H_0 is accepted. If the complex bicoherence falls outside this circle then H_0 is rejected and H_1 is accepted. Only if H_1 is accepted is the biphase test carried out.

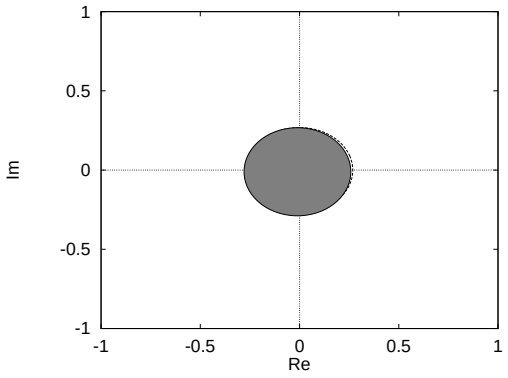

Figure 4.6: Schematic diagram showing complex bicoherence plane and magnitude acceptance region. Significant bicoherence is detected if the complex bicoherence falls *outside* the shaded area.

4.5.2.2 Testing for zero biphase

Complex bicoherence bins which are deemed to have significant bicoherence magnitude in the first test are then tested for zero biphase. The normalised biphase estimate $Z(k,l)$ (from Equation 4.8) is tested with the following hypotheses

H_0 The biphase at this bifrequency is zero (there is QPC).

H_1 The biphase at this bifrequency is non-zero (there is no QPC).

Under H_0, $Z(k,l)$ is approximately normally distributed with mean zero and unit variance[72, 87] $N(0,1)$. Thus for a given significance level $\alpha(2)$, a critical biphase

value $c_{\alpha(2)}$ can be determined from standard normal tables, such that:

$$P\left(Z(k,l) > c_{\alpha(2)}|H_0\right) = \alpha(2)/2$$
$$P\left(Z(k,l) < -c_{\alpha(2)}|H_0\right) = \alpha(2)/2 \tag{4.10}$$

where the test is now two-sided because the phase can be negative or positive. A more intuitive expression can be obtained by replacing $Z(k,l)$ with the normalised biphase (from Equations 4.8 and 4.7)

$$P\left(\theta(k,l) > \theta_c|H_0\right) = \alpha(2)/2$$
$$P\left(\theta(k,l) < -\theta_c|H_0\right) = \alpha(2)/2 \tag{4.11}$$

where the critical biphase is given by

$$\theta_c = \frac{c_{\alpha(2)}}{2K}\left(\frac{1}{b_{\text{true}}(k,l)^2} - 1\right) \tag{4.12}$$

Contours of constant probability can thus be defined combining Equations 4.11 and 4.9. As mentioned above problems arise because the variance becomes large if the true bicoherence becomes small, but for $0.05 < b^2 < 1$ the $\alpha(2) = 0.05$ acceptance region is shown in Figure 4.7. Note that as the complex bicoherence approaches the origin ($b_c \to 0$), so, from Equation 4.7, the variance of the biphase estimate, and hence the width of the acceptance region, increases without limit. Therefore the acceptance region is only shown for values of $b_c(k,l)$ down to about 0.1. It will be seen in the next section, when the two parts of the detector are considered together, that this is not a practical problem.

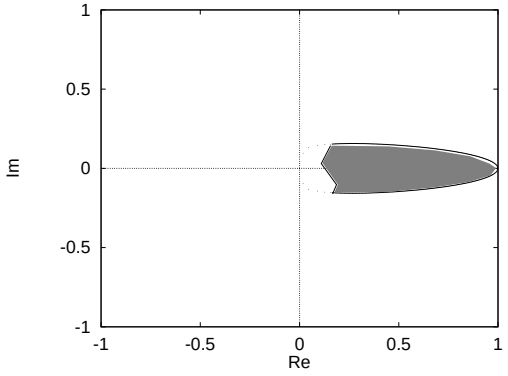

Figure 4.7: Schematic diagram showing complex bicoherence plane and phase acceptance region. Zero biphase is detected if the complex bicoherence falls within the shaded area.

4.5.2.3 QPC acceptance region

By overlaying the acceptance regions of Figures 4.6 and 4.7, it is possible to define a region of acceptance for significant QPC, as is shown in Figure 4.8. Frequency bins with a significant bicoherence magnitude fall *outside* the circle, and bins which also have a phase which is statistically zero fall within the shaded region.

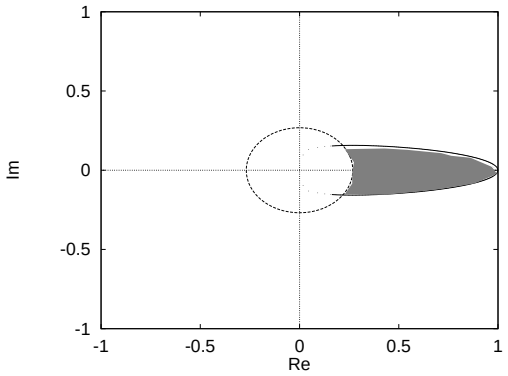

Figure 4.8: Schematic diagram showing complex bicoherence plane with QPC acceptance region. QPC is detected if the complex bicoherence falls within the shaded area.

4.5.3 Statistical properties

It is interesting to investigate, in a somewhat heuristic manner, the statistical properties of this two-part detector for QPC. The properties of any technique based on a hypothesis test can be described in terms of the probabilities of Type I and Type II errors, defined as [91]:

Type I The hypothesis is true but is rejected.

Type II The hypothesis is false but is accepted.

Because this test is a two-part test, each global verdict (i.e. there is QPC, or there is no QPC) can arise in several ways. To keep things simple, the term "hypothesis" will be applied only to the individual parts of the two part test. Five possible signal types will be considered; the M2(QPC) and M2(UC) signals already described, both with and without phase randomisation (refer back to Section 4.4 for a description of the differences between PR and CP signals) and a stochastic Gaussian signal. Firstly, the *actual* properties of such signals are summarised in Table 4.3 (which is derived from Table 4.2).

signal	significant magnitude	zero biphase	QPC ?
M2(UC)(PR)	NO	NO	NO
M2(UC)(CP)	YES	NO	NO
M2(QPC)(PR)	YES	YES	YES
M2(QPC)(CP)	YES	YES	YES
Gaussian	NO	NO	NO

Table 4.3: Summary of actual properties of signals considered in the determination of statistical properties of QPC detector.

Now Figure 4.9 shows the paths each of these signals can take through the two parts of the QPC detector, and the types of errors which can arise. It is evident that there are several possibilities of errors occuring, and it is interesting to investigate how likely these errors are to occur.

Now the problem of primary interest in this thesis is to detect QPC unambiguously in signals in which it is not known for sure whether or not the phase randomisation assumption is valid. Of interest then is the probability of false alarm P_{FA}, which is defined as the probability that QPC is detected in signals which do exhibit QPC. For each of the signal types shown in Figure 4.9 this probability P_{FA} can be written in terms of the probabilities of Type I and Type II errors. Now Type I errors are controlled by the choice of significance levels $\alpha(1)$ (for the magnitude test) and $\alpha(2)$ (for the biphase test), and these can be easily set to small values to ensure that the Type I errors are small.

Of the five signals considered in Figure 4.9, only two actually exhibit QPC, and so to focus on the P_{FA} attention is turned on the remaining three signals M2(UC)(PR), M2(UC)(CP) and the Gaussian signal.

For M2(UC)(PR) signals the P_{FA} is $P(I_{\mathrm{mag}}) \times P(II_{\phi})$, i.e. the product of the probability of a Type I error in the magnitude part of the test and the probability of a Type II error in the phase part of the test. Since $P(I_{\mathrm{mag}})$ is just $\alpha(1)$, this can be easily controlled, and so the chance of detecting QPC in an M2(UC)(PR) signal is small. A similar argument applied to the Gaussian signal.

However, the M2(UC)(CP) signal, which was the problematic signal discussed in Section 4.4, has a P_{FA} of $P(II_{\phi})$, and at the moment no indication has been given of how big this probability is. Without some understanding of where this $P(II_{\phi})$ comes

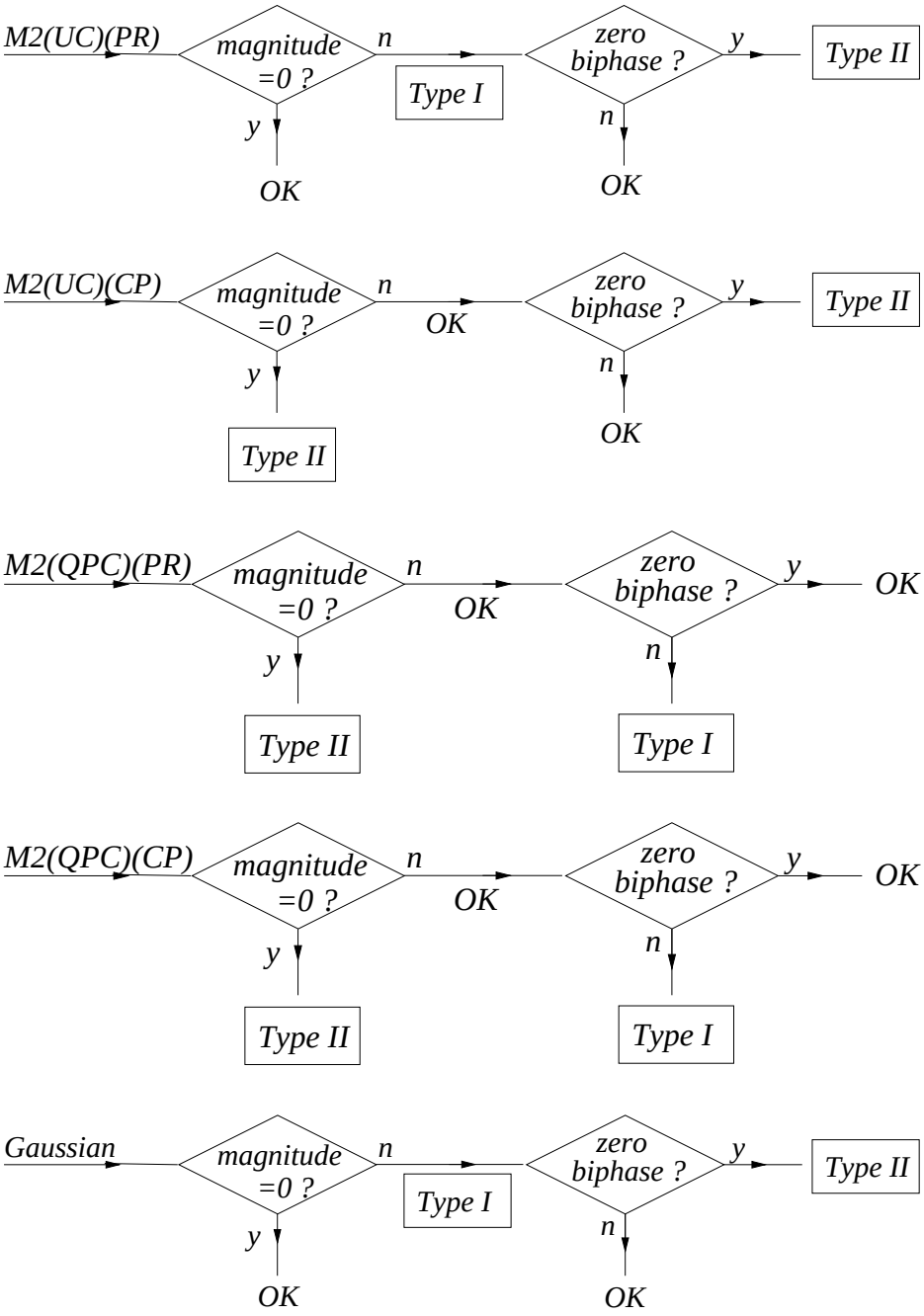

Figure 4.9: Schematic diagram showing the variety of paths different signal types can take through the two-part testing procedure, and the types of errors which arise. From top to bottom; M2(UC)(PR), M2(UC)(CP), M2(QPC)(PR), M2(QPC)(CP) and Gaussian.

from, the QPC detector has only limited applicability, since many detections might arise simply from Type II errors in the biphase test.

If there is quadratic phase coupling, it was shown in Section 4.1 that the biphase is close to zero, and, if the biphase is within a critical value θ_c (see Equation 4.11) then the zero biphase hypothesis is accepted. If there is *no* phase coupling, then the biphase will be Uniformly distributed between 0 and 2π (i.e. $\theta = U[0, 2\pi)$). Therefore, the P_{FA} is the probability that $\theta = U[0, 2\pi)$ lies within the phase acceptance region defined by θ_c, which is simply

$$\begin{aligned} P_{\text{FA}} &= \frac{\text{width of acceptance region}}{2\pi} \\ &= \frac{2\theta_c}{2\pi} = \frac{\theta_c}{\pi} \end{aligned} \tag{4.13}$$

Now making use of the approximate relation between b^2_{true} and the SNR, as given in Equation 3.18, Section 3.7.2, and derived in Appendix A.5.2, an approximate expression can be found which relates the P_{FA} to the SNR for the white noise, three sinusoid case (see Appendix B.1 for more details);

$$\boxed{P_{\text{FA}} = \frac{c_{\alpha(2)}}{2K\pi}\left(\frac{18}{M}10^{-\text{SNR}/10} + \frac{72}{M^2}10^{-2\text{SNR}/10}\right)} \tag{4.14}$$

Equation 4.14 shows that if the SNR is known (and it can be estimated from the power spectrum), then a prediction can be made of the P_{FA} of the QPC detector, and this gives a quantitative measure of how well the QPC detetcor is likely to perform in practice.

But on a note of caution, it was mentioned earlier that the expression for the biphase variance is an *approximate* one, and is likely to break down for low b^2_{true}. It follows that Equation 4.14 is likely to break down at low SNRs. This behaviour, and the quality of predictions from Equation 4.14 will be explored in some simulations which are described below.

4.5.4 Simulations

For each SNR over a wide range, one hundred signals each of the types M2(UC)(CP) and M2(QPC)(PR) were generated, and the detection rate at the discrtete bifrequency corresponding to (f_1, f_2) was measured. The signals had $N = 4096, M = 64, K = 64$. Recall that the M2(QPC)(PR) signals exhibit QPC, and so a good detector should

detect QPC for these signals, whereas the M2(UC)(CP) signals do not exhibit QPC (even though they have high bicoherence magnitude), and so a good detector will reject these.

Figure 4.10 shows a comparison between theory and experiment for these signals. Some words of explanation are given below:

M2(QPC)(PR) : signal exhibiting QPC. At high SNRs QPC is detected in the signal M2(QPC)(PR) approximately 95% of the time. This is in close agreement with the theoretical predicition, which is determined by the significance level of the biphase test, which is 5%. So the Probability of a Type I error for this signal is $\approx 5\%$. As the SNR falls the detection rate remains close to 95% until about -12dB, when the detection rate falls off rapidly.

M2(UC)(CP) : signal not exhibiting QPC. At high SNRs the detection rate for this signal is predicted (by Equation 4.14) to be ≈ 0. The simulations indicate a detection rate of about 3%, which is satisfactory agreement. As predicted, the falling SNR results in a widening in the size of the phase acceptance region, which increases the chance of signals with uncoupled phases being wrongly classified. Hence as the SNR falls the detection rate rises, although the simulation results generally give a slightly worse P_{FA} than predicted by the theory. The reason for this difference between theory and experiment is not known, but it may be due to the fact that the biphase estimator properties (i.e. that the biphase is normally distributed) are approximate.

At about -12dB, the width of the biphase-acceptance region exceeds 2π, and the detection technique fails completely. Once this happens, the P_{FA} predicted by Equation 4.14 actually exceeds unity, which is meaningless from a probabilistic point of view, and in the simulation the P_{FA} predicted by theory has been clipped to prevent this happening. Beyond this point, at still lower SNRs, the distribution of the angle which determines the width of the acceptance region tends towards a Uniform distribution between 0 and π (since all angles greater than π are wrapped back between 0 and π). As the critical biphase angle tends towards this Uniform distribution (between 0 and π), so the biphase detection decision tends towards a coin-tossing decision with $P = 0.5$. This explains why at very low SNRs the probability of detection in both QPC and UC signals tends towards 0.5.

The expressions developed in this section thus appear to be able to predict how well the

QPC detector performs. The correspondence between the theory and simulations is not exact, but the general trends are predicted well. The reason for the differences between theory and simulation are probably due to the fact that the expression for the biphase variance (Equation 4.7) is only an approximate empirical one, taken from literature, and perhaps more accurate relations can be obtained through a purely theoretical treatment. Some progress in this field by other researchers is described in Section 4.6 below, after a brief mention of the filter-invariance properties of the detector.

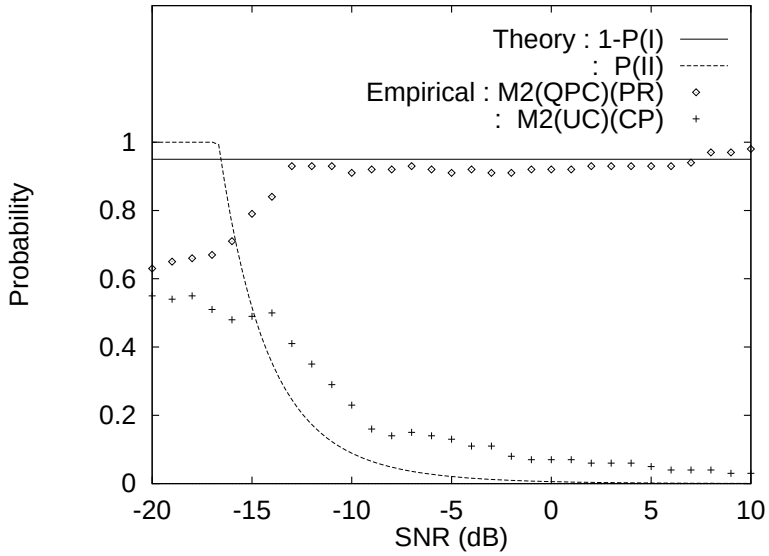

Figure 4.10: Comparison of theoretical and empirical detector performance.

4.5.4.1 Filtering

It has already been mentioned (Property 8 in Section 3.6.3) that the biphase is, in theory, unaffected by filters with linear phase characteristics. Thus, in principle at least, the performance of the QPC detector should be unaffected by such filters. Filters with nonlinear phase characteristics can greatly upset the detector performance, since they can rotate the biphase phasor outside the acceptance region, and so great care is required when designing filters which process signals before QPC detection is attempted.

4.6 Relation to Other Work

About the same time as this work was nearing completion, a similar problem was addressed by other researchers [5, 4]. Their approach is much more mathematical than the

one adopted here, but after some algebra it can be shown (see Appendix B.2,[88]) that for real sinusoids in white noise, the QPC detector proposed [5, 4] is very similar to the one adopted here. However, in their raw form, the two detectors are applied in different ways, as described in Figure B.2. In the following discussion, the detector proposed here (and in [6]) will be called the "Υ-detector" , and the other detector (described in [5, 4]) the "Ψ-detector". The main difference between the two detectors is that the Υ detector determines the variance of the biphase estimate from the bicoherence magnitude $\hat{b}^2$, whereas the Ψ detector makes estimates of the harmonic amplitudes and noise variance, and uses these to determine the biphase estimate variance. The Ψ detector is developed with no consideration whatsoever to false detections.

The Υ approach does have a flaw, since it has been shown above that $\hat{b}^2$ is not a good estimate of the true bicoherence b^2_{true} in the case where there is no phase randomisation, and also because that the expression for the biphase variance holds only for medium to high SNRs. On the other hand, the Ψ approach, which involves no averaging at all, is likely to be much more susceptible to measurement noise unless an extremely long data record is available. Furthermore the way in which the Ψ detector estimates the noise variance (by averaging all spectral bins except those which are known to contain harmonics) is likely to be biased by sidelobe behaviour, and also appears to require *a priori* information concerning the location of these harmonics.

Perhaps the best detector will emerge as a combination of the best features of each of these approaches i.e. the averaging, windowing, two-step detector strategy of the Υ detector, combined with the biphase variance estimation technique of the Ψ detector.

4.7 Summary

In this chapter the potential for using the bicoherence as a detector for quadratic phase coupling (QPC) has been examined. It has been found that under certain conditions to do with phase randomisation, the magnitude of the bicoherence can be an ambiguous measure, and a novel complex-plane representation of the problem illustrates why this is so. Recognising that the biphase does not suffer from these ambiguity problems, a new QPC detector has been developed based on the biphase *and* the bicoherence. The two-part test is able to distinguish between signals which exhibit QPC and those that do not, even if the phase randomisation assumption is breached. At high SNRs the new detector has been found to perform very well, but at low SNRs the performance falls.

Finally a connection is made between this QPC detector and another detector recently proposed elsewhere. It is argued that the best performance may be achieved with some form of hybrid detector which uses some of the good properties of the detector described here together with some of the good properties of the other one.

Chapter 5

Robust Estimation

5.1 Introduction

The squared bicoherence $b^2(k,l)$ (from Equation 3.13) can be estimated using Equation 3.14 (rewritten here for clarity);

$$\hat{b}^2(k,l) = \frac{|\frac{1}{K}\sum_{i=0}^{K-1} X_i(k)X_i(l)X_i^*(k+l)|^2}{\frac{1}{K}\sum_{i=0}^{K-1}|X_i(k)X_i(l)|^2 \frac{1}{K}\sum_{i=0}^{K-1}|X_i(k+l)|^2}. \tag{5.1}$$

For this segment-averaging approach to be valid, the signal must be stationary, in the sense that its statistical properties should not change between the K frames. However, if in fact the signal is contaminated with transients, perhaps occurring in only one or two frames, then the bicoherence estimate can be seriously distorted.

The fact that the bicoherence is sensitive to outliers resulting from transient contamination is not new information[82, 92]. Indeed the fact that HOS measures are sensitive to transient contamination has been exploited in several ways to detect transients, using the polyspectra ([93]), cumulants ([94]), or both ([95, 96]). However, although this sensitivity might be a useful property in some applications (notably underwater signal processing), it is most unhelpful in others (such as speech analysis) since transient contamination occurring in a few data segments can obscure potentially useful information occurring in all the frames.

Thus what is needed is a modification to the squared bicoherence (or skewness) estimator which renders it, in some as yet undefined way, *robust* to transient contamination.

There has already been some interest in robust techniques and HOS, interest arising naturally out of the poor variance properties of the estimators. A previous attempt to use robust estimation techniques for HOS has been concerned with third-order cumulant estimation (i.e. in the time domain) [97]. The techniques require that the triple

products used to estimate the third-order cumulants are symmetrically distributed. Unfortunately, this only happens if the signal itself is symmetrical, in which case the third-order cumulants are zero[98]. In other words, when applied to third-order cumulants, the only situations in which robust estimation techniques will work well is when there is nothing interesting to look at.

This chapter begins with the specification of a model for additive transients, which is added to the harmonics-in-noise signal model used in Chapters 3 and 4. In Section 5.2.2 an example illustrates how strongly these transients can corrupt the ordinary bicoherence estimate. To solve this problem some previous work is first described ([82]), and then several new approaches to the problem are suggested (Section 5.3). Heuristic arguments are presented for choosing between these approaches, and a new algorithm is developed which implements the best one. Finally in Section 5.4 simulations are described which assess the performance of the new algorithm in comparison with the ordinary bicoherence estimator.

5.2 The Problem

5.2.1 A model for transient contamination

Most work on QPC detection [34, 6, 5, 72] has concentrated on steady-state harmonic signals of the M2 (Equation 3.4) type. As it has already been discussed in Chapter 4, a phase randomisation assumption has to be satisfied for bispectral magnitude estimates taken from such a signal to be meaningful as QPC measures. In this chapter it will be assumed that the phase randomisation assumption is satisfied - i.e. that the sinusoid phases ϕ_1, ϕ_2 are randomized $[0, 2\pi)$ in each analysis segment.

As discussed in Chapter 4 the bicoherence of such a signal depends only on the sinusoid magnitudes A_i $i = 1, .., P$, and the magnitude of the noise $v(n)$.

In this chapter an additional transient term is *added* to $x(n)$ to give a new, more realistic

signal model;

$$\begin{aligned} x[n] &= x_{ss}[n] + x_t[n] \\ x_t[n] &= \sum_{i=0}^{N_t} t_j[n], \\ t_j[n] &= \begin{cases} 0 & n < m_j \\ T_j e^{-\alpha_j (n-m_j)} \cos[2\pi g_j (n-m_j) + \phi_j] & n \geq m_j \end{cases} \end{aligned} \tag{5.2}$$

where T_j, α_j, m_j, g_j and ϕ_j are all Uniformly distributed random variables. The transients thus have a damped sinusoidal form, similar to that used in transient analysis elsewhere (e.g. [99]).

Now the level of transient contamination can then be controlled by the range of the variables T_j, α_j and the parameter p (p is the probability of a transient occuring at any discrete time, such that $N_t \approx N \times p$). Each transient's position in the time series is determined by $m_j = U[0, N-1]$, and its amplitude, frequency, phase and decay rate by T_j, g_j, ϕ_j and α_j respectively. Using this model, signals can be generated which consist of underlying coupled sinusoids in noise, plus randomly scattered transients of random amplitudes, frequencies and phases.

In the remainder of this chapter it will be useful to have a measure of the level of transient contamination. A suitable measure is the Steady-State-to-Transient Ratio (SSTR), the ratio in dB of the variances of the transient part of the signal σ_{t}^2 to the steady-state part of the signal σ_{ss}^2 [1], so that

$$\mathrm{SSTR} \triangleq \frac{\sigma_{\mathrm{ss}}^2}{\sigma_{\mathrm{t}}^2}. \tag{5.3}$$

A signal in which the variance of the steady state component is the same as the variance of the transients will have an SSTR of 0dB. If the variance of the steady state part is higher, then the SSTR will be positive (an in the extreme case of a signal with no transients the SSTR wil be ∞dB), and if the variance of the transient component is higher then the SSTR will be negative.

[1] Note that σ_s^2 includes contributions from both the steady-state sinusoids **and** the steady-state additive noise $v(n)$.

5.2.2 Example

To illustrate the type of problem encountered, consider a simulation signal of the type given by Equation 5.2. Figure 5.1 shows the time series of three signals generated by this model with various levels of SSTR. Each signal contains three coupled sinusoids, steady state noise, and added transients. The sinusoidal components are buried in the steady state noise, so the periodicity of the signal is hidden in the waveforms shown in Figure 5.1. Figure 5.2 shows the power spectra of the same three signals, and it can be seen that the periodicty becomes visible in the spectral domain. The spectra also indicate that the transient contamination has a large effect, although the spectral peaks due to the sinusoids remain discernible even in the most heavily contaminated case. Finally, Figure 5.3 shows the squared bicoherences of the three signals.

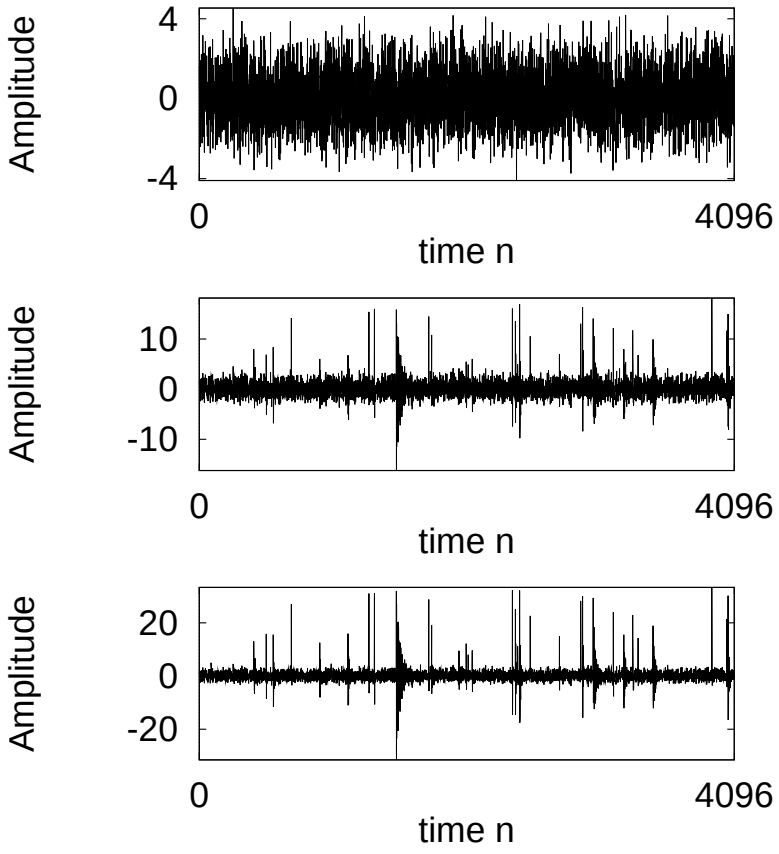

Figure 5.1: Time series of signal under different levels of transient contamination. Top : SSTR ∞ dB, middle : SSTR 0 dB, bottom : SSTR -6dB.

It is evident that the transients, of which there are only about 15 large ones, seriously distort the bicoherence picture. Although the peak due to the coupled sinusoids is still

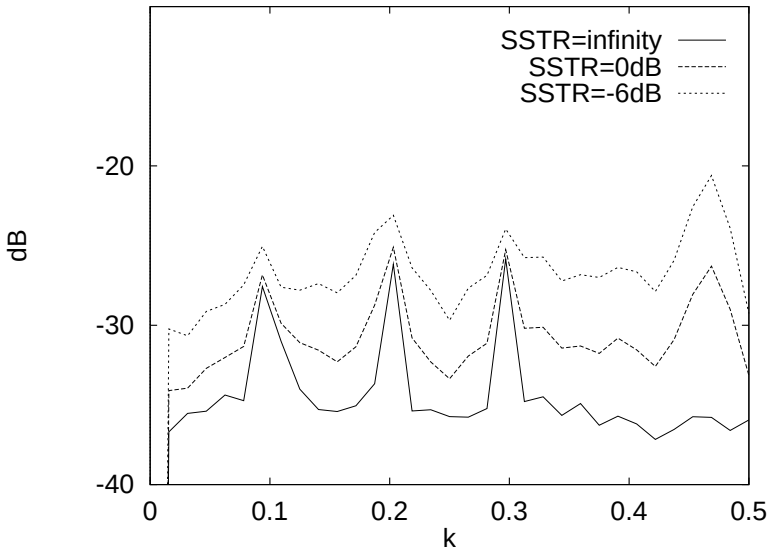

Figure 5.2: Power spectra of signal under different levels of transient contamination. SSTRs (in dB) of ∞, 0 and -6.

present at SSTR -6dB in Figure 5.3, without the prior knowledge that there is a peak at that bifrequency it is impossible to pick it out from the high background level which is due to the transients.

5.2.3 Existing Solutions

As already described in Section 3.5 Lyons et al [82] attributed the sensitivity of the bicoherence estimator to transient contamination to the fact that power spectral information (i.e. signal energy) was affecting the bicoherence estimate. To rectify this a new estimator was proposed [82] in which *all* energy information was discarded:

$$b_{\mathrm{PO}} \triangleq \frac{1}{K} \sum_{i=0}^{K-1} \frac{X_i(k) X_i(l) X_i^*(k+l)}{|X_i(k) X_i(l) X_i^*(k+l)|}. \tag{5.4}$$

Thus all segments have identical, unit weight. By rewriting the DFT in terms of magnitude and phase $X_i(k) \equiv |X_i(k)|e^{j\phi_i(k)}$ the way in which b_{PO} works becomes apparent:

$$b_{\mathrm{PO}} = \frac{1}{K} \sum_{i=0}^{K-1} \phi_i(k) + \phi_i(l) - \phi_i(k+l). \tag{5.5}$$

Equation 5.5 shows that b_{PO} is simply an average of the biphase in each frame. As such this estimator has similarities to the single segment estimate proposed in [5, 4]. The estimate gives all segments identical (unit) weight, which means that transient

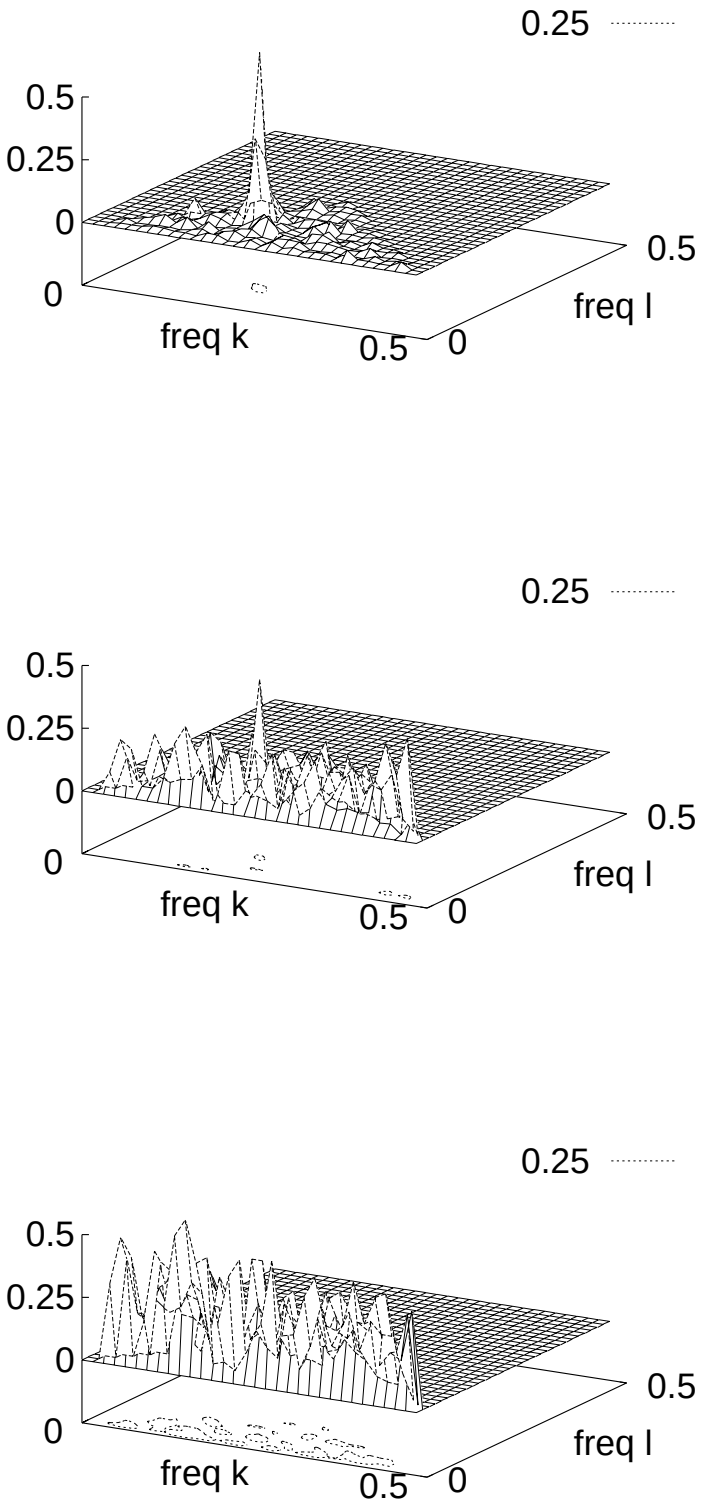

Figure 5.3: Bicoherence of signal under different levels of transient contamination. Top : SSTR ∞ dB, middle : SSTR 0 dB, bottom : SSTR -6dB.

contamination is reduced, but still remains in the estimate.

A more appealing approach is to try to remove the transients in some way so that the bicoherence estimate reflects the true underlying bicoherence. It might be possible to do this purely in the time-domain, although establishing a criteria for transient identification is somewhat difficult. Instead the approach described below aims to identify those transients which result in large bispectral values, and then aims to exclude these from the estimate.

5.3 Outlier removal

Central to the development of the approaches described in this section is the idea of viewing the bispectral samples as phasors in the complex plane, as described in Chapter 4. All the techniques described below start from modifications to the bispectral estimator

$$\hat{B}(k,l) = \frac{1}{K}\sum_{i=0}^{K-1} X_i(k)X_i(l)X_i^*(k+l). \tag{5.6}$$

The following assumptions will be made

A1 The bicoherence estimate at each bifrequency is independent of the estimate at all other bifrequencies, and so can be dealt with separately.

A2 In the complex bispectral plane, the $\hat{B}_i(k,l)$ samples which arise from segments which contain only the steady state signal components will have a different probability distribution to those samples which arise from segments containing transients.

A3 Consider the segment indices $i = 0,..,K-1$ as a set of numbers $\mathcal{K} = \{0,..,K-1\}$. Now the bispectral summation, given by

$$B(k,l) = \frac{1}{K}\sum_{i=0}^{K-1} \hat{B}_i(k,l)$$

can be rewritten as a summation over the segments indexed in $\mathcal{K}$ instead:

$$B(k,l) = \frac{1}{K}\sum_{i\in\mathcal{K}} \hat{B}_i(k,l). \tag{5.7}$$

Now view the set of all K data segments as being the combination of the segments which contain transients $\mathcal{K}_t$ and the segments which are "clean" and do not $\mathcal{K}_c$. Then the set of all segments is the union of these two sets

$$\mathcal{K} = \mathcal{K}_t \cup \mathcal{K}_c.$$

Now, defining K_c as the number of clean segments (i.e. the number of segments in $\mathcal{K}_c$), and K_t as the number of transient segments (i.e. the number of segments in $\mathcal{K}_t$), the ordinary bispectral estimate can be written

$$B(k,l) = \frac{1}{K_c + K_t}\left[\sum_{i\in\mathcal{K}_c} \hat{B}_i(k,l) + \sum_{i\in\mathcal{K}_t} \hat{B}_i(k,l)\right].$$

Now this decomposition raises the possibility of forming a better bispectral estimate by averaging only over the clean segments;

$$B'(k,l) = \frac{1}{K_c}\sum_{i\in\mathcal{K}_\rfloor} \hat{B}_i(k,l) \tag{5.8}$$

$$= \frac{1}{K - K_t}\sum_{i\in\mathcal{K}_\rfloor} \hat{B}_i(k,l). \tag{5.9}$$

Assumption A1 is a gross generalisation, and takes no account of the bispectral leakage effects described in Chapter 3. However, given the transient contamination model of Equation 5.2, this approach means that if a transient is found in a particular segment, say at bifrequency (k_1, l_1), then the contribution of that segment to the estimate at (k_1, l_1) will be removed, but the contribution at some other frequency (k_2, l_2) need not be. In this way it is hoped that data which is "clean" (unaffected by transients) will not be discarded needlessly. Of course bispectral leakage from transients will be screened out also if the algorithm developed is a powerful one.

Assumption A2 gives the view that if there is no transient contamination, then, at a particular bifrequency, the K bispectral samples will all form a cluster around the true value. Transients, which could arise from totally separate physical processes, will (it is hoped), stick out from this cluster.

Assumption A3 states that the removal of samples which do not fit in with the cluster of samples will improve the estimate. This assumption needs qualification, since in the absence of transients, the bispectral estimate over $K - K_t$ segments will in general have a higher variance than the estimate over K segments, unless $K_t = 0$. Thus, some

criterion has to be specified so that the number of transient segments K_t is not allowed to become too large.

Under these assumptions, several possible transient rejection techniques have been investigated. These are summarised in Table 5.1, and discussed in more detail below.

Algorithm	measure	computation	disadvantages
ordinary	bicoherence b^2 Equation 5.1 Section 3	low	susceptible to transients
Equal-weighting of segments	phase-only bicoherence b^2_{PO} Equation 3.17 [82]	medium	has the effect of weighting down big contributions, up-weighting small contributions
$\alpha-$trimming (trim $\Re[\hat{B}_i(k,l)]$ and $\Im[\hat{B}_i(k,l)]$ separately)	b^2_α Section 5.3.2	medium-high	normalisation problems
Iterative fixed-α outlier removal (discard $\alpha\%$ of samples)	$b^2_{\mathrm{out}(1)}$ Section 5.3.2	high	danger of discarding data samples, not transients
Iterative variable-α outlier removal (use statistics to determine when to stop outlier removal)	$b^2_{\mathrm{out}(2)}$ Section 5.3.2	high	

Table 5.1: Comparison of transient rejection techniques considered in this chapter.

5.3.1 Preliminary

As a preliminary, consider first the technique of median estimation[100]. The data vector $z(n)$ $n = 0, .., N-1$[2] are sorted and an ordered vector $Z(n)$ is formed with the properties

$$Z(0) = \min[z(n)]|_{n=0,..,N-1}$$
$$Z(N-1) = \max[z(n)]|_{n=0,..,N-1}.$$

The median value is simply $Z(N-1/2)$ (for odd N) or $\frac{1}{2}[Z(N/2-1)+Z(N/2)]$ (for even N). However, although the median is quite robust to occasional estimation errors in the $z(i)$, it is unsatisfactory because it only uses one sample value, with no averaging. The $\alpha-$trimmed mean is a generalisation of the median measure, in which $\alpha\%$ of the smallest, and $\alpha\%$ of the biggest $z(i)$ terms are excluded from the mean calculation;

[2] Note that n and N refer here to *samples* of some quantity, not necessarily the signal of interest.

5.3.2 $\alpha-$trimming

The α-trimmed mean $\overline{z}_\alpha$ is then formed by removing a proportion α of both the smallest $(Z(0), Z(1), ..)$ and the largest $(Z(N-1), Z(N-2), ..)$ values from the mean estimate. This can be viewed as a subtraction of "outliers" from the summation, or equivalently, forming a new mean estimate from a reduced set of samples classed as "inliers". These concepts will be central to the argument which follows.

The α-trimmed estimate is thus

$$\overline{z}_\alpha \stackrel{\triangle}{=} \frac{1}{N(1-2\alpha)} \sum_{i=N\alpha}^{N(1-\alpha)-1} y(i). \tag{5.10}$$

It is easy to show that if $\alpha = 0$ (i.e. there is no trimming), then $\overline{z}_\alpha \equiv \overline{z}$, the ordinary mean.

5.3.2.1 Extension to bispectrum estimation

It is possible to extend this technique to bispectrum estimation. According to A1 above the method is performed independently at each bifrequency (k, l), so these frequency indices are removed for clarity. The steps involved are:

- Treat the K segment estimates of $\Re[B_i] \equiv \Re[B_i(k,l)]\, i = 0, .., K-1$ as one set of "samples", and the estimates of $\Im[B_i]$ as another set.
- Form two ordered vectors R_i and I_i of these samples such that

$$\begin{aligned} R_0 &= \min \Re[B_i]|_{i=0,..,K-1} \\ R_{K-1} &= \max \Re[B_i]|_{i=0,..,K-1}, \end{aligned}$$

  and

$$\begin{aligned} I_0 &= \min \Im[B_i]|_{i=0,..,K-1} \\ I_{K-1} &= \max \Im[B_i]|_{i=0,..,K-1}. \end{aligned}$$

- Now compute the $\alpha-$trimmed mean estimates $\overline{R}$ and $\overline{I}$ in a way similar to that of Equation 5.10;

$$\overline{R}_\alpha \triangleq \frac{1}{K(1-2\alpha)} \sum_{i=K\alpha}^{K(1-\alpha)-1} R_i$$
$$\overline{I}_\alpha \triangleq \frac{1}{K(1-2\alpha)} \sum_{i=K\alpha}^{K(1-\alpha)-1} I_i.$$

In fact a more complicated expression could be used here [100] in which noninteger $K\alpha$ values are explicitly dealt with, but for the current discussion this simple formula will suffice. The $\alpha-$trimmed bispectrum estimate is then given by

$$\hat{B}_\alpha = R_\alpha + jI_\alpha. \tag{5.11}$$

The way in which this algorithm works, and its main disadvantage, can be understood from a consideration of the bispectral samples at one bifrequency viewed in the complex plane. Figure 5.4 shows a schematic representation of $K = 20$ bispectral samples at one bifrequency. On the left the bispectral samples are clustered around the true bispectral value, and it might be expected that the average of these points (Equation 5.6) would yield a point close to the true value. On the right part of the figure, one of the samples has been replaced with an outlier due to a transient. This transient could have a big effect on the bispectrum estimator.

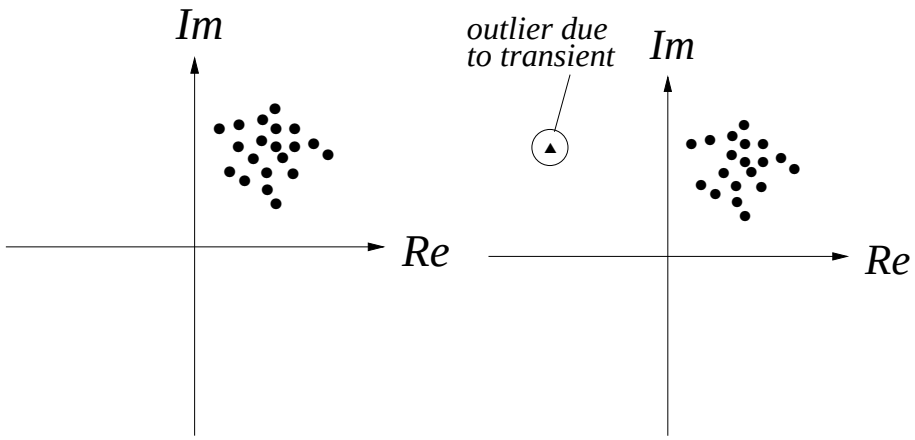

Figure 5.4: Schematic diagram illustrating the raw bispectral values from 20 segments of data. Left : clean case (i.e. no transients), right : signal with one big transient. Circles : inliers, Triangle : outlier.

Now the $\alpha-$trimming mean discards $\alpha\%$ of the largest real samples, $\alpha\%$ of the smallest real samples, $\alpha\%$ of the largest imaginary samples, and $\alpha\%$ of the smallest imaginary samples. Assume $\alpha = 5\%$, this means that one sample has to be removed from each side of the distribution of the real values, and from each side of the distribution of the imaginary values. Figure 5.5 identifies these samples.

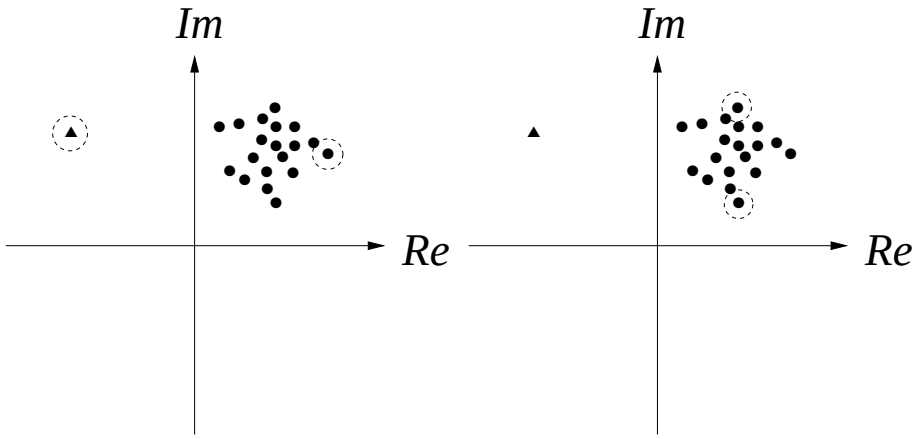

Figure 5.5: Schematic diagram illustrating the raw bispectral values from 20 segments of data. Left : samples whose real part is trimmed. Right : samples whose imaginary part is trimmed. Circles : inliers, Triangle : outlier.

In keeping with the general approach of this thesis, normalised bispectra are preferred to unnormalised bispectra, and therefore a way is sought to normalise the new bispectrum. It is here that the main disadvantage of this method arises (first reported in [101]); namely, that it is not at all clear *how* the bispectrum estimate should be normalised. It would make sense to remove the samples arising from transients from the denominator in the bispectrum normalisation, perhaps forming some sort of $\alpha-$trimmed denominator estimate in order to do so. However, if some segment has its real part trimmed, but not its imaginary part (as in Figure 5.5) , then should that data segment's contribution to the denominator be removed or included, or scaled in some way ? If separate trimming algorithms were applied to numerator and denominator, then the useful bounded property of the bicoherence would be lost too.

An alternative approach, but closely related to this one, would trim *both* complex parts, if either one of them *or* both of them are found to be outliers, and trim the denominator too. However, this would lead to the removal of somewhere between $2\alpha\%$ and $4\alpha\%$ of samples, depending on whether the real and imaginary parts which were outliers occurred in pairs (which would result in $2\alpha\%$ removal), or always separately (which would result in $4\alpha\%$ removal). Such an approach might work, but because of

this lack of even-handedness, and the lack of any underlying theory, it has not been pursued further here.

5.3.3 Iterative fixed-α outlier removal for complex outliers

A better approach is to identify outliers not by their real and imaginary parts separately, but by their *distance*, measured in the complex plane, from the centre of the sample distribution. Figure 5.6 shows how the transient sample might be identified by its distance from the centre of the underlying sample distribution (assumed circular).

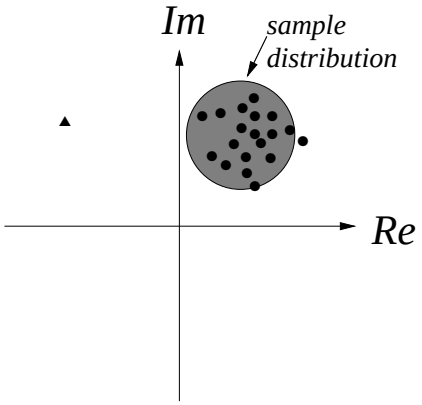

Figure 5.6: Schematic diagram illustrating the raw bispectral values from 20 segments of data. Outlier defined by distance from sample distribution. Circles : inliers, Triangle : outlier.

This method has the obvious attraction that, because outliers are identified in terms of their real and imaginary parts *together*, the difficulties described above is averted, and the contribution of transient segments can easily be removed from the denominator too, so retaining the bounded bicoherence property. In practice the underlying distribution (the shaded circle in Figure 5.6) is not known, and this has to be estimated from the data. The technique could thus be applied iteratively as follows

1. Estimate centre of underlying distribution.
2. Calculate distance of each sample from centre.
3. Reject sample furthest from centre.
4. Repeat.

However, since there is no criterion for determining whether or not the most outlying sample is indeed a transient, then there is no way of knowing *when* to stop trimming.

5.3.4 Iterative variable-α outlier removal for complex outliers

The solution to this problem can be found by making some relatively mild assumptions about the underlying distribution of the bispectral samples. These then permit the use of statistical tables to give a criterion for deciding whether the most outlying sample is an outlier or not. If it is an outlier, then it is removed and the trimming continues, but if it is not, then the trimming stops.

The key assumption which is made is that the bispectral samples $B_i(k,l)$ are samples from a multivariate (real and complex) normal distribution of unkown mean $B(k,l)$ ($\equiv$ the true bispectrum) and unkown variance. Once estimates of the mean and covariance matrix of the samples have been calculated, the distance of each sample from the centre of the distribution can be calculated.

5.3.4.1 How valid is this assumption ?

Evidence for the validity of this assumption can be found in the literature;

- In [74] it is shown that for an M1 signal (Equation 3.3) the bispectral estimate is approximately asymptotically complex normal. This means that as the data length increases, the distributions of the real and imaginary parts of the bispectrum estimate approach independent normal distributions [77]. The complex normal distribution is a special case of the multivariate normal distribution[3] in which the real and imaginary parts are independent.
- It has already been discussed in Chapter 3, and in [34, 76, 72] that b^2 estimates have χ^2 distributions. It is widely understood that the variance of the numerator of the bicoherence estimator in Equation 5.1 has a much higher variance than the denominator [34]. Therefore, it seems plausible that the statistical distribution of estimates of the squared bispectrum $|B(k,l)|^2$ will be approximately the same as those of $b^2(k,l)$. Now if $|B(k,l)|^2$ has a χ^2 distribution, it follows that $B(k,l)$ has a complex normal distribution.

[3] In which "multi" refers to the real and imaginary parts of the bispectrum.

5.3.4.2 Other changes

The distance metric can be more sophisticated than a simple Euclidean distance metric - as the sample covariance matrix is available then the Mahalanobis distance (see algorithm details below) can be used instead. This takes account for example, of the distribution having a different variance in the $\Re$ and $\Im$ directions, or even for the distribution ellipse to be rotated by some arbitrary angle, which would occur if the real and imaginary parts of the bispectum were correlated with each other. Figure 5.7 shows the familiar schematic example with an elliptical probability distribution.

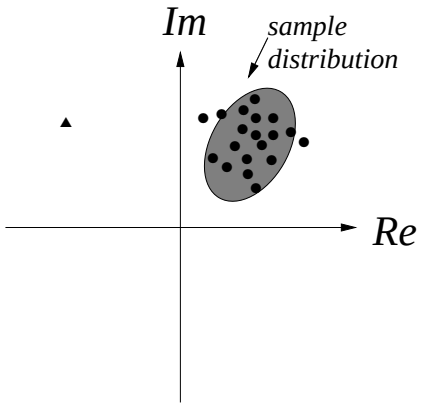

Figure 5.7: Schematic diagram illustrating the raw bispectral values from 20 segments of data. Outlier defined by Mahalanobis distance from sample distribution. Circles : inliers, Triangle : outlier.

Thus the assumption of statistical independence of the real and imaginary parts (discussed above) is not required.

Statistical tables exist which give approximate critical Mahalanobis distances for outliers from multivariate normal distributions. These can be used to test the most outlying sample, and to iteratively trim the outliers away until the most outlying sample is an inlier.

5.3.5 The Algorithm

The complete algorithm which implements these steps is described below. The computation is carried out independently for each bifrequency (k, l), and so for clarity these indices will be dropped.

1. Estimate raw bispectral estimates $B_i; i = 0, .., K-1$.

2. Initialise $K_c = K$ (i.e. assume every segment is clean).

3. (a) Form the inlying subset $B'_i; i \in \mathcal{K}_c$ of these estimates. (For the first time round the loop this contains all the data).

   (b) Estimate the mean bispectrum $\overline{B}'$ from the data.

$$\overline{B}' = \frac{1}{K_c} \sum_{i \in \mathcal{K}_c} B'_i$$

   (c) Compute the sample covariance matrix $\mathbf{C}$ of the bispectral estimate, which provides information about the orientation of the sample distribution in the complex plane;

$$\mathbf{C} = \begin{bmatrix} c_{\Re\Re} & c_{\Re\Im} \\ c_{\Im\Re} & c_{\Im\Im} \end{bmatrix}, \tag{5.12}$$

   where the covariance terms are given by[4]

$$\begin{aligned}
c_{\Re\Re} &= \frac{1}{K_c - 1} \sum_{i \in \mathcal{K}_c} (\Re[B'_i - \overline{B}'])(\Re[B'_i - \overline{B}']), \\
c_{\Re\Im} &= \frac{1}{K_c - 1} \sum_{i \in \mathcal{K}_c} (\Re[B'_i - \overline{B}'])(\Im[B'_i - \overline{B}']), \\
c_{\Im\Im} &= \frac{1}{K_c - 1} \sum_{i \in \mathcal{K}_c} (\Im[B'_i - \overline{B}'])(\Im[B'_i - \overline{B}']), \\
c_{\Im\Re} &= \frac{1}{K_c - 1} \sum_{i \in \mathcal{K}_c} (\Im[B'_i - \overline{B}'])(\Re[B'_i - \overline{B}']).
\end{aligned}$$

   (d) Identify the sample which is separated from the sample mean by the largest Mahalanobis distance;

$$D^2_{\max} = \max_{i \in \mathcal{K}_c} (B'_i - \overline{B}')^T \mathbf{C}^{-1} (B'_i - \overline{B}').$$

   (e) Under the assumption that the underlying distribution is multivariate normal, approximate critical values of $D^2_{\max}$ are available in statistical tables [102, Table XXXII]. Compare $D^2_{\max}$ with the tabulated threshold values for the required significance level, for example T_5 (at the 5% level).

   (f) If $D^2_{\max} > T_5$ then the sample corresponding to $D^2_{\max}$ is classed an outlier, and is excluded from future calculations. K_c is decremented by one.

[4] It is evident that $c_{\Re\Re}$ and $c_{\Im\Im}$ can be rewritten as simple squares.

(g) Go back to (a) until $D^2_{\max} <= T_5$ (so that the outermost sample is classed as an inlier) or until some maximum number of samples have been discarded.

4. The modified squared bicoherence function is then recomputed from Equation 5.1 with each term replaced by its modified primed term, where $\sum \equiv \sum_{i \in \mathcal{K}_c}$ now, and all averages are taken over the K_c inlying segments $i \in \mathcal{K}_c$.

$$\hat{b}^2_{\text{trim}}(k,l) = \frac{|\frac{1}{K_c}\sum_{i\in\mathcal{K}_c} B_i(k,l)|^2}{\frac{1}{K_c}\sum_{i\in\mathcal{K}_c}|Z_i(k,l)|^2 \frac{1}{K_c}\sum_{i\in\mathcal{K}_c}|P_i(k+l)|^2}. \tag{5.13}$$

5.4 Simulations

To see how well the algorithm described above performs, several simulation experiments have been carried out. These all use the signal model of Equation 5.2 with a variety of input parameters controlling the level of transient contamination. The fact that there are such a large number of parameters in the model makes it difficult to explore every avenue, and so a subset of the possible models has been studied.

As a first look, the examples described in Section 5.2.2 will now be revisited with the new estimation algorithm. Figure 5.8 shows the new estimator for the same signal with three SSTR conditions. Comparing Figures 5.3 and 5.8 it is immediately apparent that a great improvement has occurred. The effect of the transients has effectively been screened out, and the new estimator clearly shows the QPC peak.

To gain further insight into the performance of the detector, multiple simulations have been carried out. For each of these, Equation 5.2 is used, with sinusoids at frequencies f_1, f_2 and $f_1 + f_2$. Two new measures are defined for measuring the effectiveness of the detector at QPC detection, these are;

r_1 the (empirical) probability that the highest point in the bicoherence occurs at the discrete bifrequency associated with (f_1, f_2). For N_s simulations, this is defined as

$$r_1 \triangleq \frac{\text{Number of times peak occurs at } (f_1, f_2)}{N_s} \tag{5.14}$$

r_2 the average proportion of bicoherence "energy" which occurs at the discrete bifre-

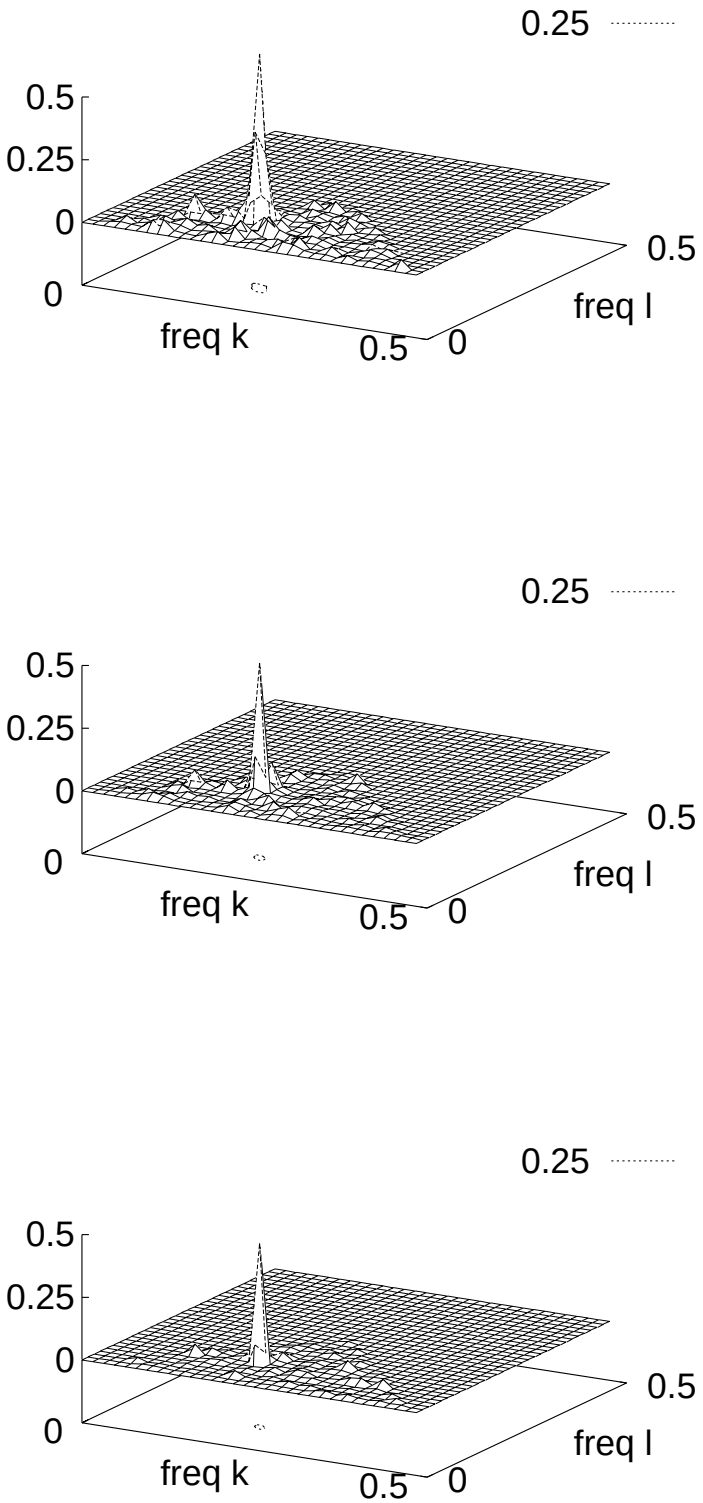

Figure 5.8: Bicoherence of signal under different levels of transient contamination, using new robust estimator. Top : SSTR ∞ dB, middle : SSTR 0 dB, bottom : SSTR -6dB.

quency associated with (f_1, f_2), defined as

$$r_2 \triangleq \frac{\sum^{N_s} b^2_{\text{trim}}(f_1, f_2)}{\sum^{N_s} \sum_{IT} b^2_{\text{trim}}(k, l)} \tag{5.15}$$

Both r_1 and r_2 are bounded between 0 and 1, and good QPC detectors will have higher values of each measure than poor detectors.

Figure 5.9 shows the values of r_1 determined from $N_s = 100$ simulations, over a range of SSTR values. For these simulations the level of steady state background noise (Gaussian) to the steady-state sinusoids was such that, when no transients were added, the SNR was 0dB. The LHS of Figure 5.9 corresponds to a signal with no transients, while the RHS corresponds to a signal contaminated with large transients. Simulations were performed for SSTRs between 12dB and -18dB, at 3dB intervals, and in addition the extreme LHS of this figure shows the signal with no transients, in which the SSTR is ∞dB. Thus the $x-$axis of this plot is nonlinear at the leftmost point.

Starting at the LHS, with the SSTR = ∞ dB, both detectors correctly identify the peak at (f_1, f_2) 100% of the time. As the tranisents' amplitudes are increased (SSTR falls), the orindary bicoherence detector begins to detect peaks at other bifrequencies, and so r_1 falls. The new detector however, successfully screens these transients out, and retains a high detection rate. Ultimately, when the SSTR becomes very low, the new detector begins to fail too. This is due to the fact that as the SSTR falls the estimated covariance matrix $\mathbf{C}$ (Equation 5.12) becomes more and more distorted by the transients and so the radius of the estimated distribution increases, so it becomes more difficult for the algorithm to distinguish between outliers and inliers.

Figure 5.10 shows the values of r_2 determined from the same $N_s = 100$ simulations, with SSTRs between 12dB and 18dB at 3dB intervals, plus the ∞dB case for no transients. In this case the mean and standard deviation of r_2 are shown. At infinite SSTR, the proportion of bicoherence energy at (f_1, f_2) is high for both detectors. Interestingly r_2 is slightly higher for the ordinary detector than for the new detector. This is because the algorothm sometimes mistakenly identifies samples as outliers, and discards them, resulting in a small degradation in performance (of a few %) for the case of signals which are not contaminated by transients.

As the SSTR falls, the performance of the ordinary detector starts to fall, as the transients result in new bicoherence peaks occurring at other bifrequencies. The new detector, however, is again successful in ignoring these transients, resulting in an ap-

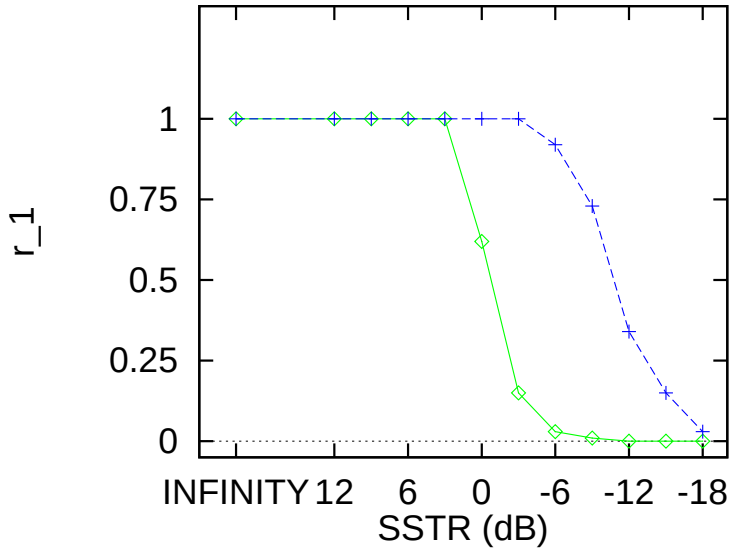

Figure 5.9: r_1 as a function of SSTR, obtained from 100 simulations. (Note the SSTR-scale is nonlinear at the left hand side). Diamonds : ordinary estimator, Crosses : new robust estimator b^2_{trim}.

proximately constant r_2. At very low SSTR's the new detector begins to fail, for the same reasons as described above.

5.5 Summary

In this chapter the issue of transient contamination of bicoherence estimates has been investigated. It has been shown that transients occurring in a small number of data segments can have a large influence over the direct-method bicoherence estimate. If these transients are not wanted then this poses a real practical problem. After a consideration of several interesting aspects of this problem, a novel solution has been proposed which is based on a statistical multivariate outlier rejection test carried out in the complex bispectral domain. Some simulation results indicate that the new estimator is able to exclude interfering transients succesfully, but that if the transients become very large this method fails too.

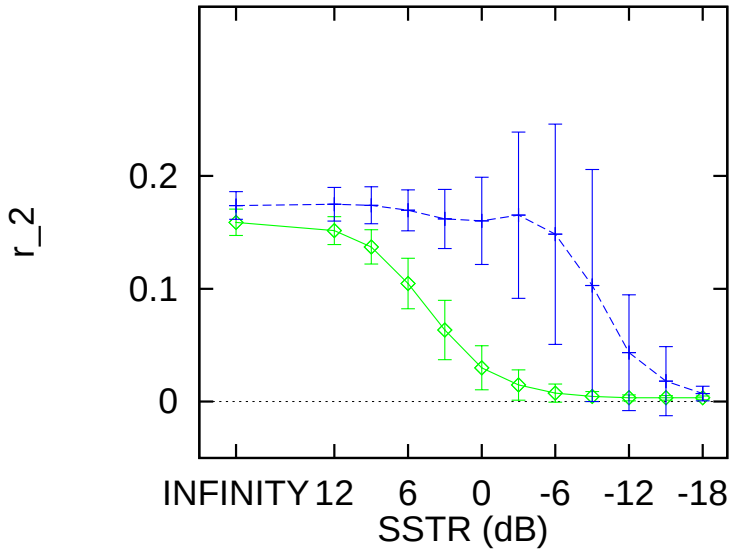

Figure 5.10: Means (points) and standard deviations (error bars) of r_2 as a function of SSTR, obtained from 100 simulations. (Note the x-scale is nonlinear at the left hand side). Diamonds : ordinary estimator, Crosses : new robust estimator.

Chapter 6

Experimental Technique

The fundamental questions which form the motivation for the bispectral analysis of speech signals are:

- Is there evidence of quadratic nonlinearities in speech sounds ?
- Can HOS-based features reliably characterise speech sounds ?
- Can HOS-based features reliably characterise speakers ?
- How are the HOS-based features related to each other ?
- How are the HOS-based features related to conventional features ?

This chapter discusses why existing speech databases are not suitable for bispectral speech analysis, and then describes a speech database assembled especially for nonlinear and HOS analysis. The complete specification of the database, and details of how the speech was recorded are given in Appendices C.1 and C.2.

Each speech sound in the database can be described by a variety of measures, including conventional measures (such as power spectra and spectral moments) as well as new HOS measures (such as bicoherence and QPC detection matrix), and the remainder of the chapter describes the choice of parameters for analysis using these measures. In Chapter 7 these measures will be used to analyse the speech sounds in the database.

6.1 Preliminary: Phonetics

The discussion of speech sounds in Chapter 2 did not use any formal framework for determining what constitutes "a speech sound". Since such a formalism is in widespread use, and will be of use in this and the following chapter, the idea of phonemes and notation for their representation, will now be briefly described, although there are many introductory texts on this subject available [1, 2, 8].

A *phoneme* is defined [2] as "the smallest unit in speech where substitution of one unit for another might make a distinction of meaning". For example, the vowels in the spoken words "had" and "hid" are sufficient to distinguish between their meaning, and so have different phonemes. (Note that the fact that the words are spelled differently is not the issue here - phonemes are defined in terms of the **sounds** of the spoken words).

Several notation schemes have been used in the past, but the two which will be used (concurrently) in this thesis are the IPA (International Phonetic Alphabet) [1] and the SAM-PA (Speech Assessment Methodology Phonetic Alphabet) [12]. In IPA phoneme symbols are enclosed in oblique lines (e.g. /i/), but individual realisations of sounds are enclosed in square brackets (e.g. [i]). For SAM-PA phoneme symbols are shown in typewriter font (e.g. `i`). (Table C.1 in Appendix C.1.1 shows symbols from both alphabets, together with words which contain these phonemes.)

6.2 A Speech Database for HOS Analysis

Most HOS analysis of speech to date has been carried out on very small data sets, with sometimes just one speaker, a few sounds, and no repetitions of each sound (e.g. [44, 36]). Inferences made about the HOS properties of speech signals from such small databases are unlikely to be meaningful, since natural speaker-to-speaker, sound-to-sound and utterance-to-utterance variability is ignored. Therefore it seems wise to use a database in which there are multiple speakers, multiple speech sounds, and multiple repetitions of each sound.

As an added complication, the fact that HOS estimators have high variances means that reliable results cannot be obtained from short data records. For the purpose of spectral analysis, speech is usually regarded as stationary, or at least quasi-stationary, over a window of length 10-30ms[103]. While this window is long enough for reliable spectral estimates to be formed, it is generally not long enough for reliable bispectral estimates[1], a problem that will now be explained.

In a paper concerned with stochastic signals [67], it has been suggested that reliable bispectral estimates can be obtained if the total data length N is more than the square of the DFT size M (since [67] suggests that $K \geq M$, and with no overlap $N = KM$)[2].

[1] Unless of course the frequency resolution of the estimate is greatly reduced.

[2] These criteria are for *stochastic signals*, such as unvoiced fricatives, but for signals with *determinis-*

Now the frequency resolution Δf of the final bispectrum estimate will be f_s/M, and so the data length T (in s) required for reliable estimates, according to the guideline from [74], is related to the specified frequency resolution and the sample rate by the following relation:

$$T = \frac{N}{f_s} = \frac{M^2}{f_s} = \frac{f_s}{\Delta f^2}$$

Figure 6.1 shows how this required data length (plotted here in ms) varies with frequency resolution for three possible sample rates. It is evident that, even for a bispectrum with rather coarse frequency resolution, a very long data record is required, with the worst case plotted being a high resolution $\Delta f = 50$Hz bispectrum, which would require nearly 10 seconds of data at the highest sampling rate.

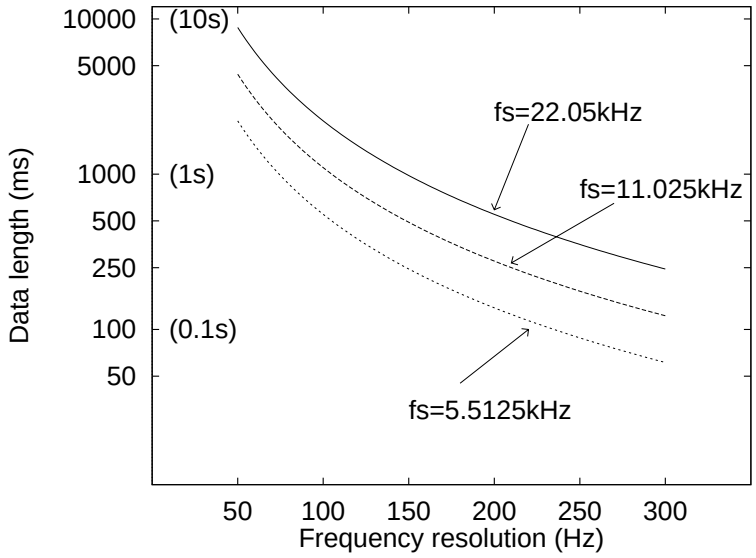

Figure 6.1: Relation between frequency resolution Δf and data record length T (here in ms) for reliable bispectrum estimation for three sampling rates.

Thus it seems that the requirements of the bispectral estimator for reasonable estimates are totally at odds with the quasi-stationary data lengths available from continuous speech in existing speech databases.

One solution to this problem is to collect multiple occurrences of each sound, and then to carry out ensemble averaging over these (as done by [29, 28] for speech and by [104] for drum sounds), but this ignores the natural utterance-to-utterance variability in the signal, which for speech sounds might itself hold interesting information. For this reason the current work does not use this approach.

tic components, the situation is somewhat different. This matter is covered in some detail in Section 6.3, and the data lengths required for voiced speech sounds are subsequently calculated in Appendix C.4.

An alternative way forward would be to develop refined HOS methods which require less data. In power spectral analysis *parametric* estimators are capable of lower variances than nonparametric estimators, and the same is true for bispectral estimators. However, the disadvantage of parametric methods is that they require modelling assumptions (e.g. that the process is AR), and there is a danger that the assumptions could obscure the interesting, and unkown, HOS properties of the signal of interest. For this reason parametric techniques are also excluded from further consideration in the current study.

A third solution is to find some unusually long utterances of the sounds of interest, and carry out the bispectral analysis on these. This can be done by assembling a database in which the speakers are asked to hold each speech sound for much longer than normal. In fact this type of approach is popular in speech pathology [44, 45], but for different reasons : laryngeal problems often manifest themselves more clearly over long data records.

In the current work this third approach, using lengthened speech sounds, is preferred, since it still allows for utterance-to-utterance variation, it requires relatively few assumptions about the speech production mechanism, and it should ensure that the bispectral estimates are reliable. This last point is of particular importance, because previous work [105] has illustrated that, unless the data lengths used are known to be sufficient for reliable results, HOS estimates can be very misleading.

It should be realised that in choosing this approach an assumption is being made - that the underlying physical processes which occur in lengthened speech sounds are the same as those which occur in continuous, normal speech. It should be realised that lengthened speech sounds contain none of the effects of coarticulation, effects which are extremely important (and difficult to model) in fluent speech. In defence of this assumption is the fact that lengthened speech sounds do have very similar spectral characteristics to sounds taken from continuous speech, and that both lengthened and normal speech sounds are produced in a very similar way. The idea underpinning this approach is that *if* interesting HOS properties are found in a database of lengthened speech, *then* it will be worthwhile developing new estimation techniques to work with shorter data lengths.

The composition of the database, and the details of how it was recorded, are given in Appendix C. The completed database consists of speech and Laryngograph records of 16 speakers (10M, 6F), saying each of 23 test words 5 times each. The 23 test words can be divided into vowels (12), fricatives (voiced 4, unvoiced 4) and nasals (3).

Before considering the analysis strategy adopted in detail, the idea of pitch-synchronous analysis will be introduced, since this technique is needed to get unambiguous bicoherence estimates from the voiced speech signals.

6.3 Pitch-Synchronous (PS) Analysis

The simple segment-averaging approach described in Section 3.4 is well suited to signals which are primarily stochastic in nature, but if the signal has strong periodic components then problems in interpretation can arise. Of course power spectral estimates can be formed using a segment-averaging approach for such signals, and the periodicity of the signal manifests itself as a periodicity in the power spectrum - this effect is well understood [84]. It can be shown [105, 89, 90] by a consideration of this effect in the polyspectral domain, that the bispectrum and bicoherence of a strongly periodic signal is itself periodic *in both frequency directions k and l*, that is, that the bicoherence of a strongly periodic signal resembles a "Bed of Nails" [89, 90]. The peaks in the power spectrum of such a signal are separated with frequency f_0, the fundamental of the signal. In a similar way the peaks (or "nails") in the bicoherence are separated by f_0. While this is an interesting phenomenon, its occurrence does not necessarily indicate anything interesting about QPC[3], even though it is sometimes taken as an indicator of this [28].

Furthermore, it has been found that if the fundamental frequency f_0 of the signal changes, then the periodic structure disappears more rapidly from the bispectrum (and bicoherence) than from the power spectrum. This effect, which has been observed in simulations and experiment [105] indicates that simple segment-averaging bicoherence estimation approaches may not be appropriate for a signal which has a varying fundamental frequency. Now the fundamental frequency of voiced speech, determined by the vibration rate of the glottis, is known to vary a great deal, and this suggests that it may be difficult to get reliable bicoherence estimates if the pitch varies over the recording window.

A way of avoiding this problem is to carry out the analysis *pitch-synchronously*. This involves a modification of the simple segment-averaging approach described in section 3.4 to extract M data points from each pitch cycle. The technique is described in detail in

[3] because of the Phase Randomisation issues discussed in Chapter 4.

Appendix C.4.

6.3.1 QPC detection for voiced sounds

For the purposes of bispectral analysis, voiced speech can be considered as a *determinstic signal in noise* estimation problem. Because each data segment is synchronised with a triggering event (the glottal closure from the Laryngograph signal)[4], the *phase* of any frequency component present in the speech will be *the same* for each speech frame. Thus the PS frames not only have the same spectral content, they also have the same phases. This can be verified by seeing that the waveforms from two consecutive pitch cycles *look* the same[5]. As a consequence of this, the phase randomisation assumption discussed in Chapter 4 cannot be made for PS voiced speech, and it is known (see Section 4.4) that if phase randomisation is *not* a valid assumption, then the squared bicoherence for such signals does *not* necessarily indicate the presence of QPC. In light of this fact, that the squared bicoherence of voiced speech sounds is likely to be an ambiguous measure of QPC, the two-part QPC detector developed in Chapter 4 must be used for voiced speech analysis.

6.4 Analysis Strategy

Suitable parameters for the bispectral analysis of speech sounds can be chosen by referring to results described in Chapters 3 and 4. The guideline $N \geq M^2$, suggested in [67], will be followed, since this provides some measure of confidence in the results. Furthermore, the Hamming window will be used in all estimates for the speech sounds, because in Section 3.7.1 it was shown that this is a good choice of data window for the detection of peaks in the squared bicoherence.

Other parameter choices are dictated by the properties of the speech signals themselves. For example practically all the useful spectral information for most speech sounds occurs at frequencies below 5kHz, and so it is often permissable to subsample the data (recorded at 22.05kHz) down to 11.025 kHz prior to analysis.

For each speech sound, a subset of the database is used for a detailed preliminary

[4] The effects of noise are neglected for the moment.

[5] This is because the waveform encapsulates phase *and* magnitude information.

analysis, to get some idea of what sort of results to expect for the conventional and HOS-based features. This preliminary analysis is confined to the speech of one speaker, *jw*. Of particular interest in each case is how the features vary over the 5 utterances of each sound, as well as how the features vary from sound to sound.

In addition to HOS-based features, a number of conventional speech measurement features are also calculated for each speech utterance. The following sections describe the various steps taken in the analysis of the database.

6.4.1 Preprocessing

Although the application of pre-emphasis to speech signals is a common practice, it is not used in the analysis described here. Consequently the power spectra of the speech sounds exhibit a negative spectral slope. Property 6 in Section 3.6.3 indicates that linear pre-emphasis should not have any effect on the bicoherence magnitude, and as long as the pre-emphasis filter has a linear phase characteristic it should not affect the QPC detector either. In the light of this information, there appears to be little purpose in applying pre-emphasis to speech signals prior to HOS analysis. One area in which analog pre-emphasis might be useful is in boosting the signal at high frequencies to control quantisation noise, but this avenue has not been explored in the current work.

For each utterance 40000 samples are recorded at $f_s = 22.05$kHz, and subsequently the voiced speech data is subsampled to 11.025kHz. This has the advantage that the subsampled signal has significant spectral content across the frequency range 0-5kHz, and so the bicoherence plots of the subsampled signal are *not* mostly empty.

The subsampling is achieved as follows; first the raw signal is filtered with a crude low-pass filter to give the filtered signal $x'(n)$:

$$x'(n) = \frac{1}{2}x(n) + \frac{1}{2}x(n+1).$$

Then every other sample of $x'(n)$ is discarded to yield the subsampled signal. The filter is a crude one, and it has only a gentle roll-off. However, this is unlikely to lead to significant aliasing problems because the speech signals are *not* pre-emphasised, and so the energy levels of the speech signals near the folding frequency are small, and this means that if any aliasing does occur, then it will be at a very low level. The fact that this filter has a linear phase characteristic is important from a QPC detection point of

view, as mentioned in Section 4.5.4.1.

Fricative sounds can have nonzero spectral content at higher frequencies (> 5kHz) [8], and so the subsampling procedure may not always be ideal for fricatives. In the analysis presented in Chapter 7 the preliminary direct comparison between voiced and unvoiced fricative speech for one speaker uses subsampled data, and subsequent analysis uses subsampled data for the voiced fricatives, but full bandwidth (f_s = 22.05kHz) data for the unvoiced fricatives. The reason for this is to focus attention on the potentially interesting low-frequency information in the voiced sounds, but to allow the possibility of high frequency information in the unvoiced sounds.

The record lengths required for these analyses are as follows; for unvoiced sounds a minimum data length of 4096/11025 = 0.37s for the preliminary analysis, and 4096/22050 = 0.19s for the comprehensive analysis; for voiced sounds the length of original speech required depends on the fundamental frequency because of the PS approach (The relation between these quantities is described in Appendix C.4.3).

For each utterance only 4096 samples were used, with a DFT size M of 64 with no overlap. This results in a frequency resolution of 22050/64 = 345Hz for the unvoiced fricatives and 11025/64 = 172Hz for the subsampled signals. This is a coarse resolution, larger than the bandwidths of many formants, and so it may not be fine enough to pick out some fine structure in the signals of interest. However, to increase the resolution would require significantly longer data records, which for speech sounds would be unfeasible. The philosophy behind this approach is that it is sensible to first see if there is anything interesting in the HOS of the speech sounds using these coarse tools, and to develop more intricate tools only if the coarse analysis yields promising results.

6.4.2 HOS Measures

For the voiced speech sounds, the analysis is applied pitch-synchronously, and for the unvoiced speech sounds a normal segment-averaging approach is sufficient. In order to be able to compare multiple bicoherence plots, a simplified form of plotting the bicoherence is adopted, as shown in Figure 6.2. Only the IT is drawn for these plots, and rather than show the squared bicoherence magnitude in 3-d, a contour plot is used. Note that only one contour level is drawn in the contour plot, and that even though most of the plot appears empty, this is because much the squared bicoherence magnitude is *above* and not below, the contour level.

The main purpose of looking at such plots is to find out whether there is significant bicoherence magnitude, since for voiced sounds this information is used in the first part of the QPC test (the second part being the test for zero biphase), and for unvoiced sounds this is on its own enough evidence for QPC. As discussed in Appendix A.2, under the null hypothesis of "Gaussianity", each sample bin of b^2 is approximately $\chi^2/2K$ distributed with 2 degrees of freedom[6]. The contours for this plot are drawn at the level $c^{\chi^2}_{\alpha(1)}/2K$. Thus bifrequency bins which have $\hat{b}^2$ above the contour level are significant at the $\alpha(1)$ level. The choice of $\alpha(1)$ is somewhat arbitrary, but from experience a very small $\alpha(1)$ has been found to give good results with simulation signals, so in the speech results in this thesis, $\alpha(1) = 0.001$, so $\chi^2_{\alpha(1)} = 13.8$. The sum of the squared

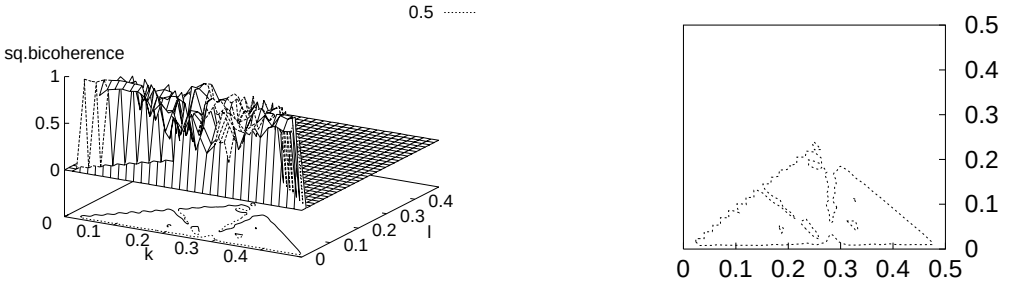

Figure 6.2: Two ways of displaying the squared bicoherence.

bicoherence over all the IT $\sum_{\text{IT}} b^2$ gives a summary measure of the squared bicoherence level in the IT. A closely related summary statistic is $\overline{b^2_{\text{IT}}} \triangleq \sum_{\text{IT}} b^2$, the average of the squared bicoherence over the IT. This is useful because it can be interpreted in two ways; from a nonlinearity detection perspective as an indicator of the average extent of QPC (for PR signals, see Chapter 4); and from a "Gaussianity" test perspective as an indicator of how non-Gaussian the signal is (since $\overline{b^2_{\text{IT}}}$ is closely related to the summary test statistic for "Gaussianity" $\sum_{\text{IT}} s^2$ proposed by [74] as shown in Appendix A.2.3).

For voiced speech sounds, the biphase is tested for zero at bifrequency bins in which the squared bicoherence is found to be significant. If the biphase is found to be within the critical values (defined by another significance level $\alpha(2)$), then a QPC detection is declared. By trial and error a suitable value for $\alpha(2)$ has been found to be 0.05.

[6]The indices (k, l) are dropped for clarity

Figure 6.3 shows two ways of displaying the QPC detections - in each plot the bifrequencies in which QPC has been detected are given the value 1, and where it is not detected, 0. This function is denoted $q(k,l)$. The sum of the QPC detections over all the IT $\sum_{\text{IT}} q$ gives a summary measure of the number of QPC detections in the IT.

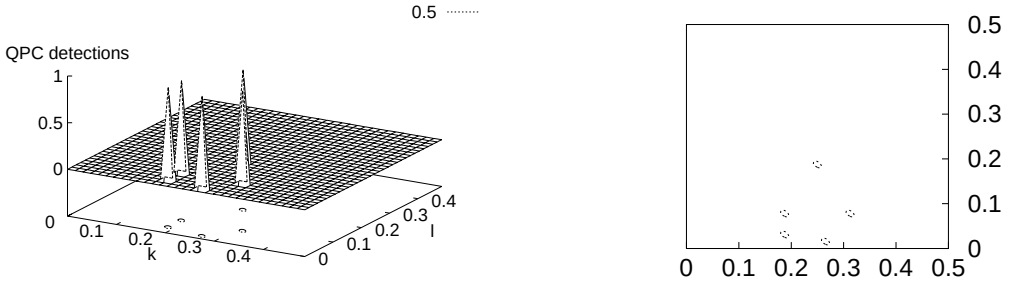

Figure 6.3: Two ways of displaying QPC detections.

For each of the speech sounds the skewness function $s^2(k,l)$ is also computed. In Section 3.5 a comment was made that the squared bicoherence and skewness function are often very similar. By calculating these quantities for each speech sound the correlation between these features can be measured. The skewness function has been proposed as a measure of "non-Gaussianity" (in the IT [74]) as well as for stationarity (in the OT [67]), and so, even though these topics are outside the main focus of interest (nonlinearity detection), they are calculated for comparison with the other measures. In addition the summations of the skewness function in the IT and OT have been proposed as components in a detector for speech activity [54], so these features could be useful in assessing the validity of that approach.

6.4.2.1 Subzones of the IT

Since the summations over the IT or OT described above are just single numbers, they do not provide much detail about what the shape of the actual quantity of interest. In other words, they are *coarse* measures. To obtain more detailed information about the IT content, the IT can be subdivided into four subzones, as shown in Figure 6.4. The

subzones have different shapes, but have approximately equal area[7], and the content of each subzone is distinguished from the others primarily in terms of the highest frequency component included, which is determined by the diagonal line $k+l = constant$. For example, if one signal contains only low frequency components, then its squared bicoherence[8] average in subzones 3 and 4 should be low. The average of b^2 over the L_i bifrequencies in zone i will be denoted $\overline{b_i^2} = \frac{1}{L_i}\sum_{z_i} b^2$, and a similar notation is used for q and s^2.

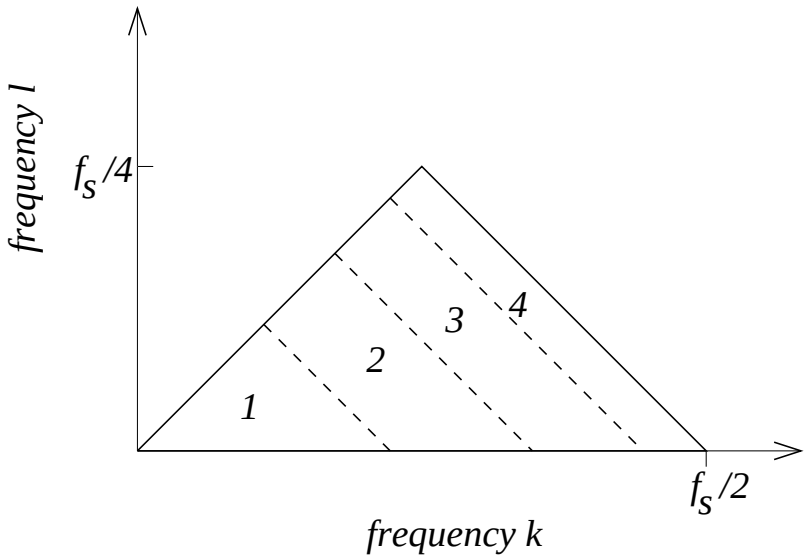

Figure 6.4: Technique of dividing the IT into four sub-zones.

6.4.3 Conventional Measures

In addition to the HOS measures described above, a number of conventional measures are also computed to see if they are correlated in any way with the HOS measures. These conventional measures are based on the power spectrum, and give some information about the energy distribution of the signal. For the vowel sounds, the formant frequencies are commonly used to differentiate between different vowel sounds [8]. The formants are calculated using an LPC model in the Entropic Waves speech analysis package. For the fricatives (both voiced and unvoiced) the spectral moments[9] appear to be reliable measures[106], and so these too are calculated. It should be stressed that these spectral moments are **not** HOS measures, but are based on a rather ad-hoc method of characterising the shape of the power spectrum. Previously published

[7] This ensures that the variances of the subzone means are approximately equal.

[8] or skewness or QPC detection plot.

[9] The computation of the spectral moments is described in Appendix C.3.

fricative analyses used the Hamming window[106], and so it is used here too.

6.5 Summary

In this chapter the speech database has been described, and a set of measurements - some from HOS and some from conventional techniques - have been selected for use in the analysis. These features are summarised in Table 6.1.

Measure	Summation region (if applicable)	symbol	sound class(es)
Squared bicoherence	Summed over IT	$\sum_{\mathrm{IT}} b^2$	all
	Averaged over IT	$\overline{b^2_{\mathrm{IT}}} \stackrel{\triangle}{=} \frac{1}{L} \sum_{\mathrm{IT}} b^2$	all
	Averaged over z_i $i = 1, .., 4$	$\overline{b^2_i} \stackrel{\triangle}{=} \frac{1}{L_i} \sum_{z_i} b^2$	all
QPC detections	Summed over IT	$\sum_{\mathrm{IT}} q$	all
	Averaged over IT	$\overline{q_{\mathrm{IT}}} \stackrel{\triangle}{=} \frac{1}{L} \sum_{\mathrm{IT}} q$	all
	Averaged over z_i $i = 1, .., 4$	$\overline{q_i} \stackrel{\triangle}{=} \frac{1}{L_i} \sum_{z_i} q$	all
Skewness function	Summed over IT	$\sum_{\mathrm{IT}} s^2$	all
	Summed over OT	$\sum_{\mathrm{OT}} s^2$	all
	Averaged over z_i $i = 1, .., 4$	$\overline{s^2_i} \stackrel{\triangle}{=} \frac{1}{L_i} \sum_{z_i} s^2$	all
Spectral moments	-	m_1, m_2, m_3, m_4	fricatives
Formants	-	f_0, f_1, f_2	vowels

Table 6.1: Features used to characterise speech database. The (k, l) indices have been omitted from $b^2(k, l)$, $q(k, l)$ and $s^2(k, l)$ and quantities derived from them for clarity.

Chapter 7

Bispectral Analysis of Speech Database

7.1 Introduction

This chapter describes the properties of the speech database in terms of the HOS measures described in Chapters 3 to 5. The primary aims are to establish whether or not speech sounds have significant levels of squared bicoherence, whether they exhibit QPC, and whether there are any patterns in the HOS measures which may be useful in future research.

The results are organised as follows : Each of the three speech classes - fricatives, vowels and nasals, is treated separately in Sections 7.2 to 7.4. For each class, some preliminary results are presented from analysis of the speech of just one speaker (*jw*): This preliminary stage allows identification of key areas of interest, and will prove helpful in the context of the analysis of the database as a whole. After the preliminary step, the properties of the speech for all the speakers are considered. Each set of results is then discussed, to try to interpret what it is that is being measured in each case. In Section 7.5 the results of correlation analysis are presented which give some indication of how the HOS-based and conventional features are related to one another. The application of the robust estimation algorithm developed in Chapter 5 is covered in Section 7.6, and the feasibility of further analysis is discussed in Section 7.7.

7.2 Fricatives

7.2.1 A First Look

Figure 7.1 shows the complete 40000-point records for eight fricatives (4 unvoiced, 4 voiced) spoken by speaker *jw*. From these time series it is evident that

- The time series appear to be relatively steady-state. This is encouraging because the segment-averaging approach used for bicoherence estimation relies on the stationarity of the signal. Any large amplitude modulation effects, or discontinuities in the time series would be indicators that the estimator may not work.
- There is a large negative DC bias on the recordings. This is not a problem in practice, since the mean is subtracted from each segment during estimation.
- There is considerable amplitude variation *between* sounds. Since the bicoherence is a dimensionless measure, it is not dependent on absolute signal amplitude. However, given that the background noise level is likely to be approximately constant, the signals with lower amplitudes will have a lower SNR. In this experiment the background noise level is very low, since the recordings were made under studio conditions, and so the between-sounds variation in amplitude should not affect the results adversely.
- The utterance `Z` appears to end after about 35000 samples (at $f_s = 22.05\text{kHz}$)[1]. This type of problem occurs in several of the speech sounds recorded, but it can be shown from considerations in Appendix C.4.3 that this data length is more than enough for the required DFT size.[2] The data length required for voiced speech analysis is indicated on the time series for `Z` in Figure 7.1.

Zooming in to the first 4096 samples (Figure 7.2) of these time series reveals the stochastic structure of the unvoiced fricatives and the periodic structure of the voiced fricatives, although the voiced fricatives still contain a visible stochastic component.

[1] This corresponds to 17500 samples of the subsampled signal with $f_s = 11.025\text{kHz}$

[2] After subsampling $f_s = 11.025\text{kHz}$, and with $K = M = 64$, the minimum data length for unvoiced sounds (see Section 6.4.1) is $N = 4096$ (at 11.025kHz), while for voiced sounds (see Appendix C.4.3), the miniumum data length is 0.64s, which corresponds to $N = 7000$ (at 11.025kHz) or $N = 14000$ (at 22.05kHz).

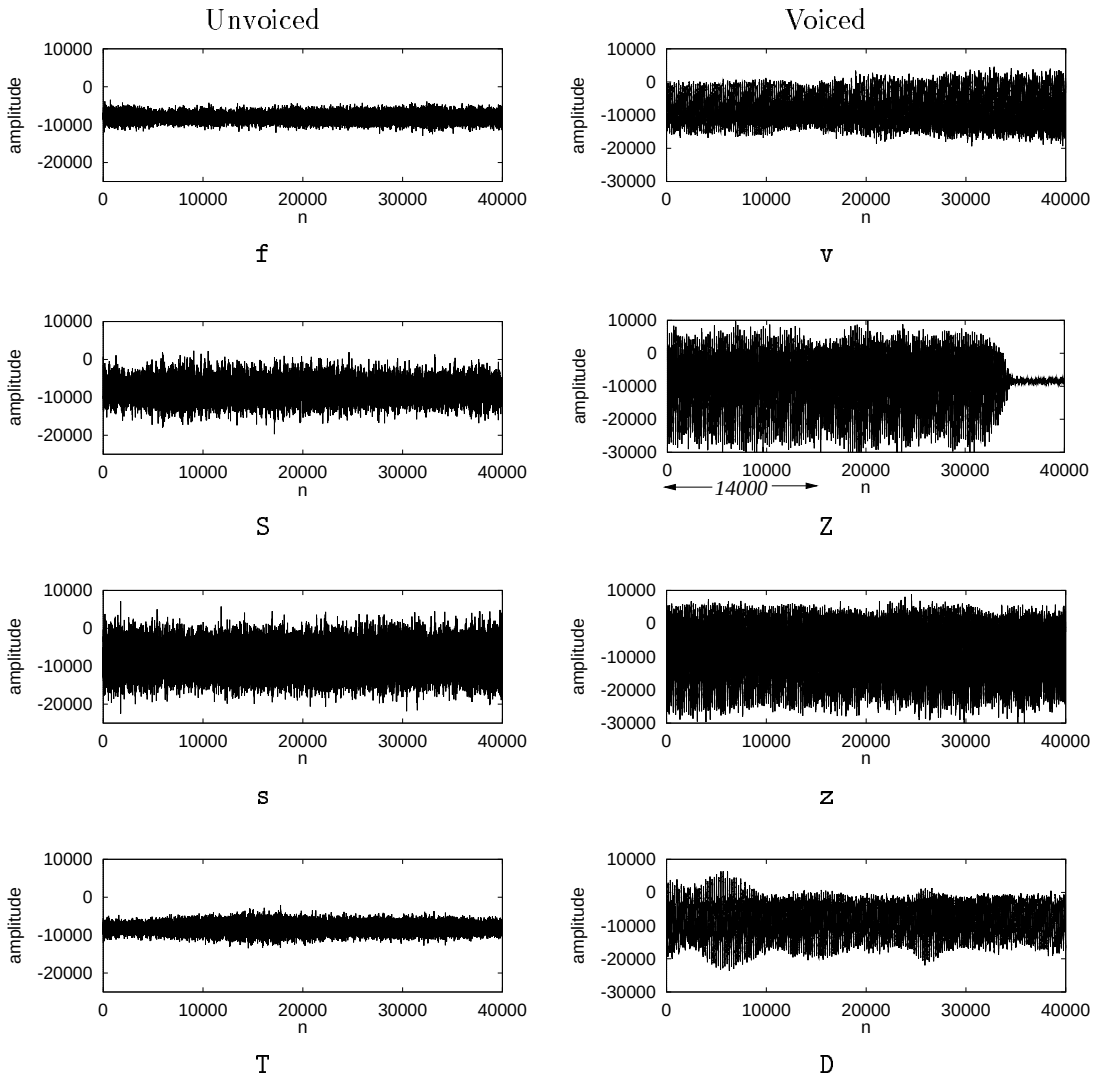

Figure 7.1: Complete time series of the eight fricatives (one utterance of each sound) spoken by *jw*. $f_s = 22.05$kHz. The time series of Z shows the approximate size (≈ 14000 samples) of the frame over which bispectral analysis is performed.

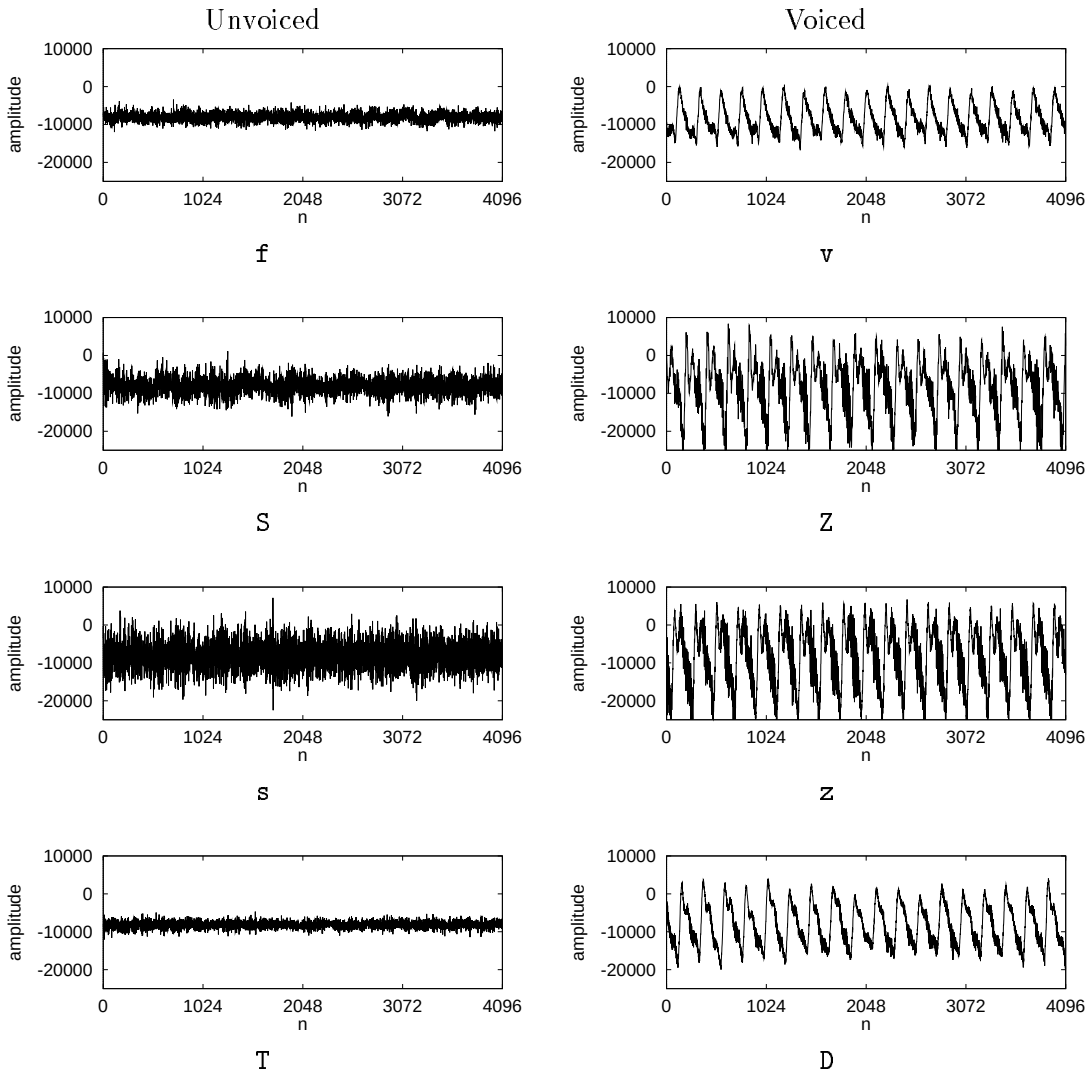

Figure 7.2: Zoomed time series (first 4096 points) of the eight fricatives (one utterance of each sound) spoken by *jw*. $f_s = 22.05$kHz.

Further insight can be obtained from the power spectra of these sounds, shown in Figure 7.3. These power spectra are computed from subsampled signals ($f_s = 11.025$kHz) with a DFT size of $M = 64$, $K = 64$ segments, and a Hamming data window. The spectra of the voiced sounds are computed pitch-synchronously as discussed in Section 6.3.

In each of these plots the power spectra of all 5 recorded utterances for each word are shown. It is evident that there is a fair amount of consistency across utterances of any one sound, although differences seem to be larger at high frequencies than at low frequencies. Furthermore it can also be seen that the spectra of the voiced sounds (on the right hand side of Figure 7.3) bear a slight resemblance to those of the corresponding unvoiced sounds, with added spectral content at low frequencies. For example the spectra of the voiced fricative `z` (as in "buzz") resemble the spectra of the unvoiced fricative `s` (as in "bus"), with a positive spectral slope in the region 0.2-0.4f_s ($\approx 2 - 4$kHz). An exact correspondence between the stochastic parts of the spectra is not to be expected, however [8], since the unvoiced sounds are produced with one point of excitation (where the fricative noise is produced) while for voiced sounds the VT is also excited by the glottis.

The fact that the power spectra for the 5 repetitions are close to each other indicates that the power spectral estimator used here has quite a low variance. It is difficult to quantify exactly how low the variance of the estimator is, because there is an (unknown) amount of natural variation between multiple realisations of the same sound. Nevertheless it appears that the spectral measure is quite a reliable one.

The squared bicoherences of the same eight sounds are shown in Figures 7.4 (unvoiced fricatives) and 7.5 (voiced fricatives). Following the discussion in Section 6.4.2 (and Appendix A.2.1) the contours are at 0.1, which corresponds to the $\alpha(1) = 0.001$ level for significant squared bicoherence magnitude at each bifrequency. As it was discussed in Section 4.5 the squared bicoherence can also be interpreted as a measure of the proportion of energy at any one bifrequency which is quadratically phase coupled. The contour levels therefore indicate which bifrequencies exhibit more than 10% of QPC.

The striking thing about the squared bicoherences of both unvoiced and voiced sounds is that they are all very low, with very few bifrequencies having significant levels of squared bicoherence magnitude. The bifrequencies which do have significant magnitude are quite evenly spread across the IT for the unvoiced fricatives, but seem to be clustered around the low bifrequency area for the voiced fricatives.

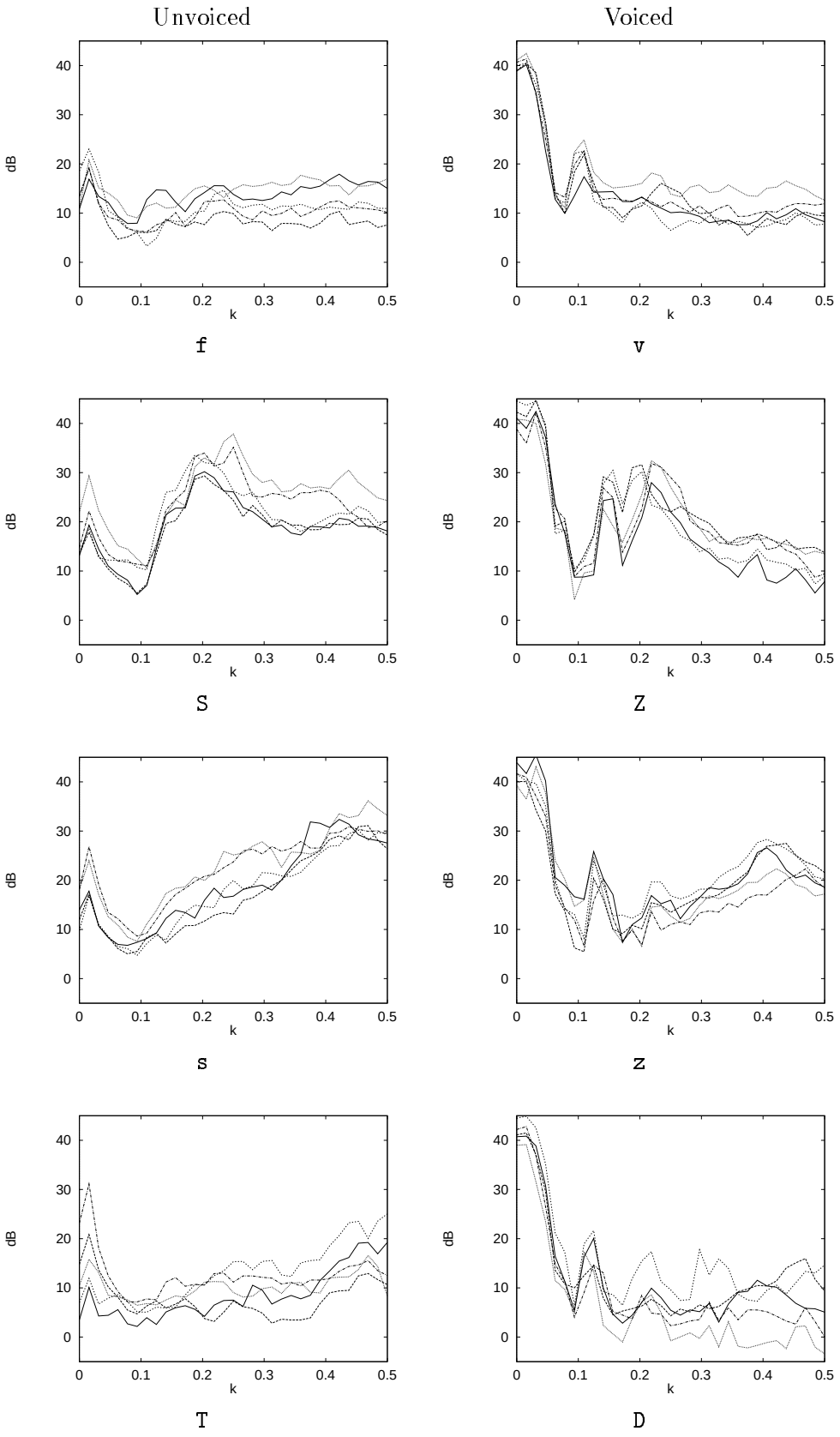

Figure 7.3: Power spectra of the four test sounds spoken by *jw*. Each plot shows the power spectrum of each of the 5 recorded utterances. Frequencies shown are normalised. $f_s = 11.025$kHz, $N = 4096$, $M = 64$, Hamming window.

Beyond this observation there appear to be no obvious patterns in the magnitudes of the squared bicoherence which distinguish one sound from another. Indeed, in contrast to the spectral measures described above, there seems to be a general lack of consistency in the bicoherence plots across utterances.

Figures 7.6 and 7.7 show the QPC detection plots for the same unvoiced and voiced fricatives. The unvoiced fricatives (Figure 7.6) show very low levels of QPC, with several utterances yielding no QPC detections at all. The number of detections in these plots is very small, and it is possible that the number of detections which *are* made are due not to QPC, but to Type II errors in the QPC detector. More will be said about this matter in Section 7.2.3.2.

The voiced fricatives (Figure 7.7) have higher numbers of QPC detections, and like the b^2 plots in Figure 7.5 these are again clustered toward the lower left hand corner of the bifrequency plane. Although this could be taken as an indication of QPC detection, there again seems to be no discernible pattern in the QPC detections, and the number of QPC detections, and the bifrequencies at which these detections occur, change greatly from utterance to utterance. The fact that the bifrequencies at which QPC detections are made seems to change so much from utterance to utterance suggests that these QPC detections *may* also be Type II errors.

7.2.2 Descriptive Statistics

Further information can be obtained by seeing how the speech features described above vary across the whole database. Figures 7.8 and 7.9 show scatter diagrams of $\overline{b^2_{\mathrm{IT}}}$, the squared bicoherence averaged over the IT[3], for the unvoiced and voiced fricatives of all speakers. The analysis parameters for the voiced sounds are as described in the section above (with subsampled data, $f_s = 11.025$kHz), but the analysis of the unvoiced sounds is carried out without subsampling, as described in Section 6.4.1.

There is a scatter diagram for each sound (classified according to SAM-PA notation), and each point on each diagram represents the value of $\overline{b^2_{\mathrm{IT}}}$ for one recorded sound. For example, it can be seen from the bottom diagram in Figure 7.8 that there are only three instances of `T` spoken by speaker *kh* (for reasons discussed in Section C.1.4).

These diagrams are very informative, and deserve close scrutiny. From Appendix A.2.3

[3] See Section 6.4.2 for a definition of this feature.

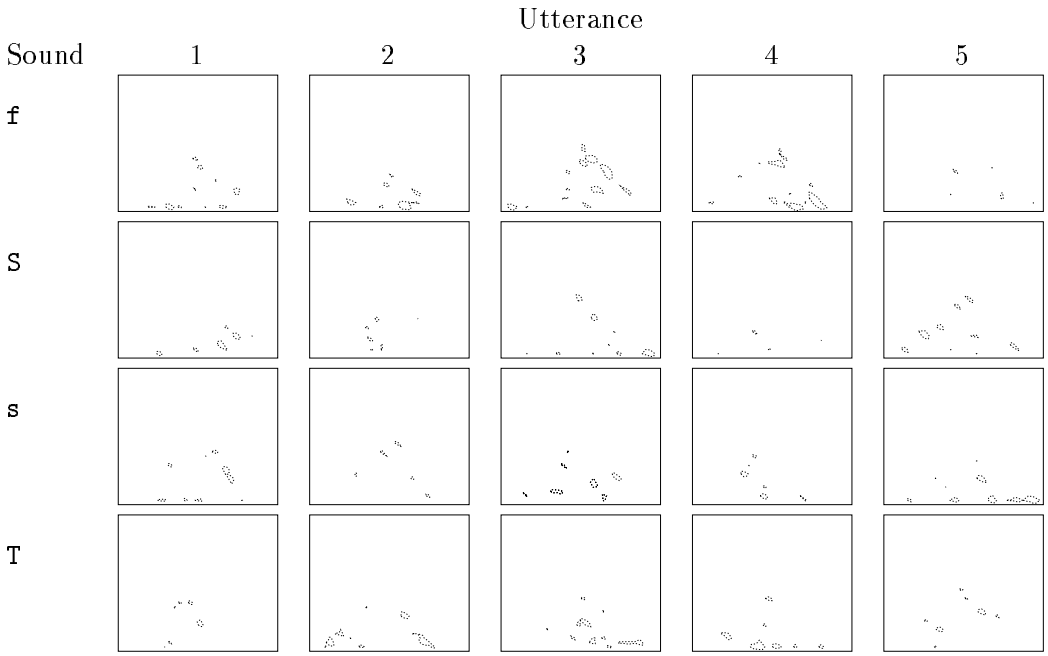

Figure 7.4: Squared bicoherences for 5 utterances of 4 unvoiced fricatives for speaker *jw*. For each plot, the two frequency axes range from 0-5.5125 kHz ($f_s = 11.025$kHz), only the IT is shown, and the contours are at 0.1, which is the $\alpha(1) = 0.001$ significance level for squared bicoherence.

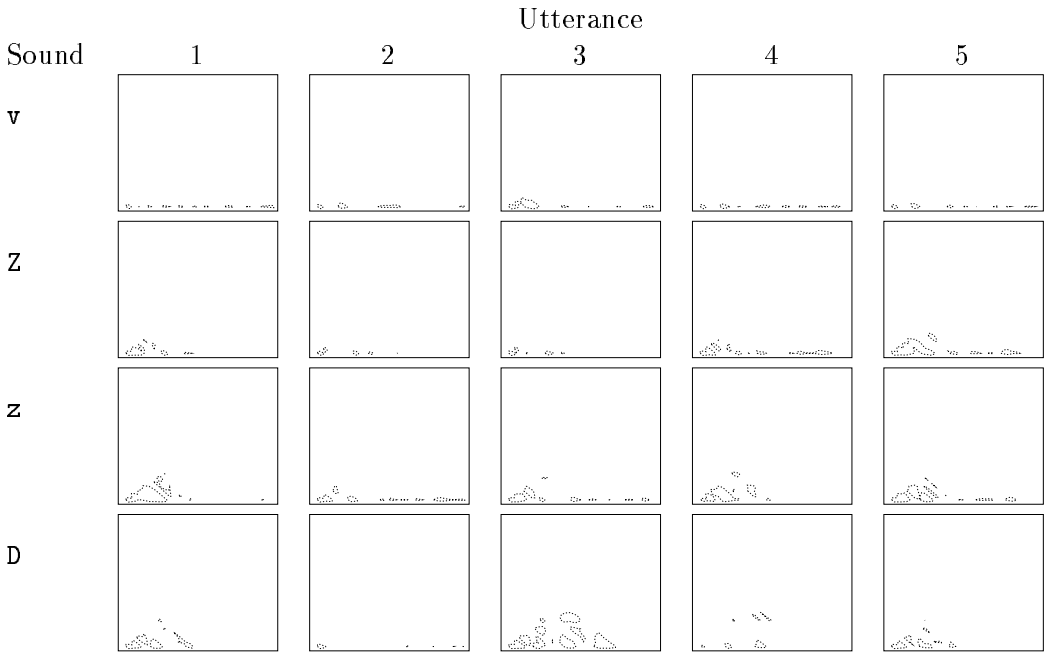

Figure 7.5: Squared bicoherences for 5 utterances of 4 voiced fricatives for speaker *jw*. For each plot, the two frequency axes range from 0-5.5125 kHz ($f_s = 11.025$kHz), only the IT is shown, and the contours are at 0.1, which is the $\alpha(1) = 0.001$ significance level for squared bicoherence.

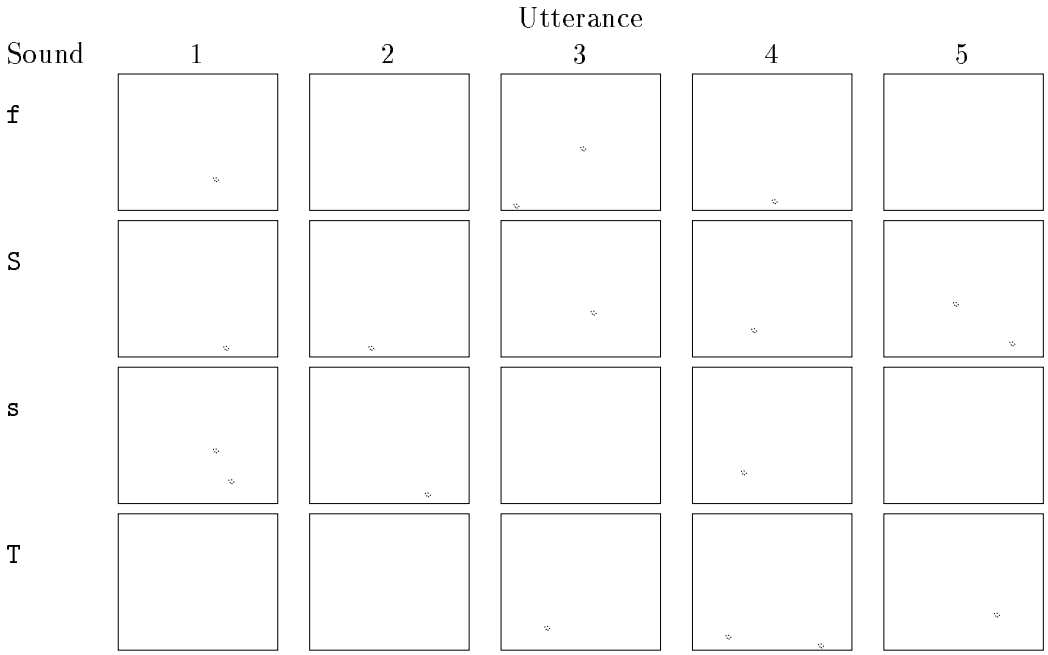

Figure 7.6: QPC detections for 5 utterances of 4 unvoiced fricatives for speaker *jw*. $f_s = 11.025$kHz.

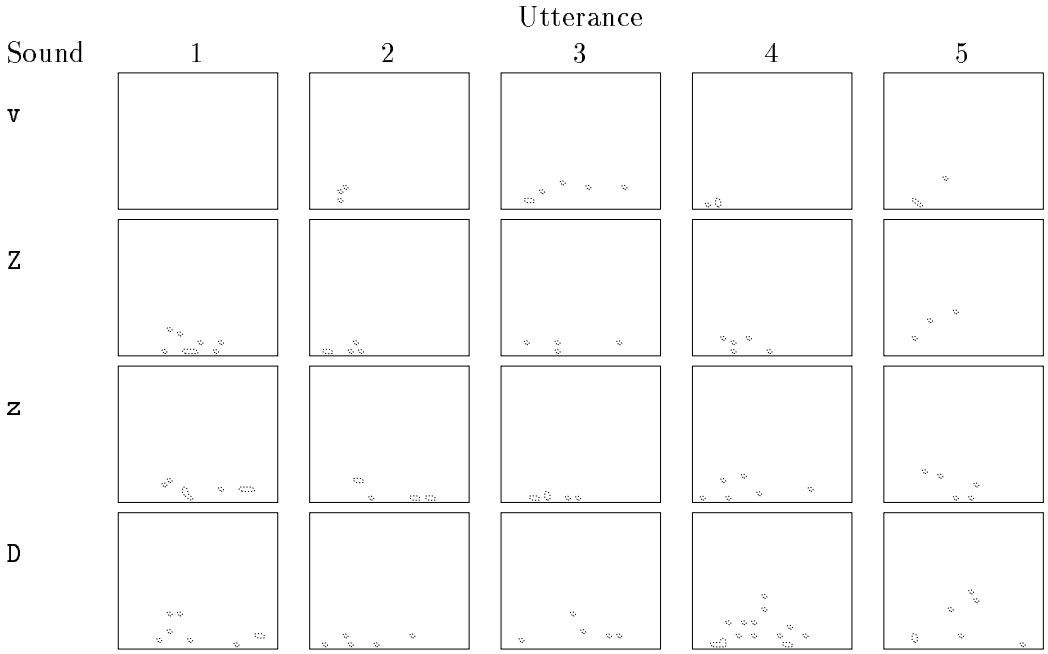

Figure 7.7: QPC detections for 5 utterances of 4 voiced fricatives for speaker *jw*. $f_s = 11.025$kHz.

the threshold for significant $\overline{b^2_{\rm IT}}$ for $K = 64$ at the $\alpha(1) = .001$ level is $c = 0.019$, and this threshold is shown in the scatter plots. The values of $\overline{b^2_{\rm IT}}$ for the unvoiced fricatives of very many of the speakers are below this threshold value. Following Appendix A.2 this could be taken as evidence of the "Gaussianity" of the unvoiced fricatives.

Looking closely at Figure 7.8 it is apparent that one speaker, *jf*, shows the highest variability for all four sounds. At the time of the recording it was noted that speaker *jf* was suffering from a cold, but in the absence of any control data it is not at all possible to infer anything about the importance of this factor.

The scatter plots of the voiced fricative data (Figure 7.9) are very interesting, since there are large differences in the values of $\overline{b^2_{\rm IT}}$ for different speakers. Many of the utterances recorded have bicoherence magnitudes which exceed the 0.1 threshold, but some speakers (*ca*, *cs*, *eb*, *kh*, *rw* and *ta*) have average squared bicoherences close to unity, whilst others (*gu*, *is*, *jf*, *jw*, *lb*, *mc*, *md*, *ps*, *pb* and *sw*) have levels close to zero. What is perhaps surprising is that the speakers in this first group (with high $\overline{b^2_{\rm IT}}$) are *all female*, whilst the second group are *all male*. Thus it appears that the squared bicoherence responds very differently to female speakers compared to male speakers. One possible explanation for this is that there is more high-frequency energy in female speech than male speech, and that as a result of this the females' speech has structure at the high frequencies, whereas the males' speech becomes noise dominated as the frequency increases.

Figures 7.10 and 7.11 show the percentage of bifrequency bins in the IT (out of a total of 225) in which QPC is detected. Evidently there are very few QPC detections for any utterance from any speaker. The low level of the squared bicoherence magnitude for unvoiced fricatives results in a very low QPC detection rate (in Figure 7.10). The situation for voiced fricatives is different, since many more of these sounds, especially those spoken by females, had significant b^2 at many bifrequencies. Figure 7.11 shows that the proportion of bifrequencies at which QPC is detected is still generally less than 10%.

Although it was observed earlier that there is no discernible pattern in the QPC detections for the speaker *jw*, there are some trends in these scatter plots which are interesting. The most interesting of these is that the female speakers (*ca*, *cs*, *eb*, *kh*, *rw* and *ta*), who all had far higher levels of $\overline{b^2_{\rm IT}}$ than the male speakers, have very low rates for QPC detections, much lower in fact than the male speakers. Thus although the bicoherence magnitudes are more significant for female speakers than for male speakers,

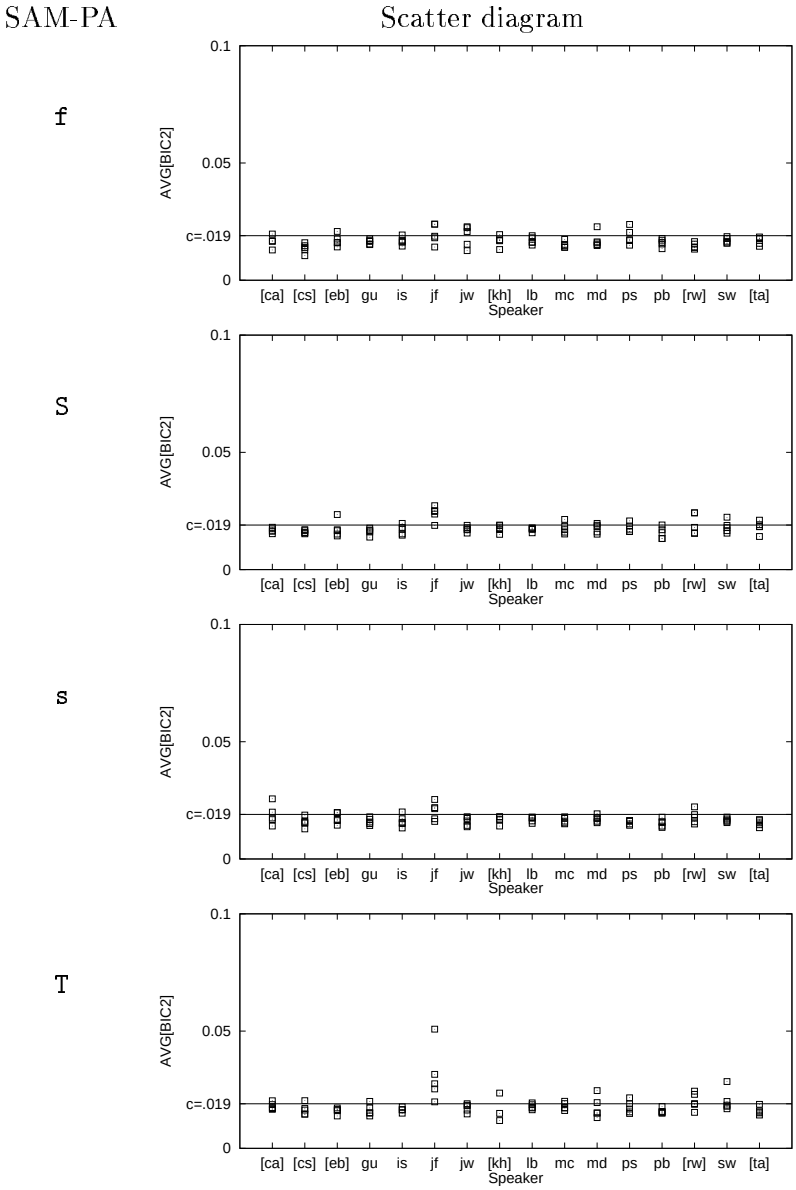

Figure 7.8: Scatter diagrams showing $\overline{b^2_{\mathrm{IT}}}$ for unvoiced fricatives by all 16 speakers. $f_s = 22.05$kHz. The null hypothesis H_0 that the signal is Gaussian is rejected at the $\alpha = 0.001$ level if $\overline{b^2_{\mathrm{IT}}} > c$ the critical level. Female speakers are denoted by [ ].

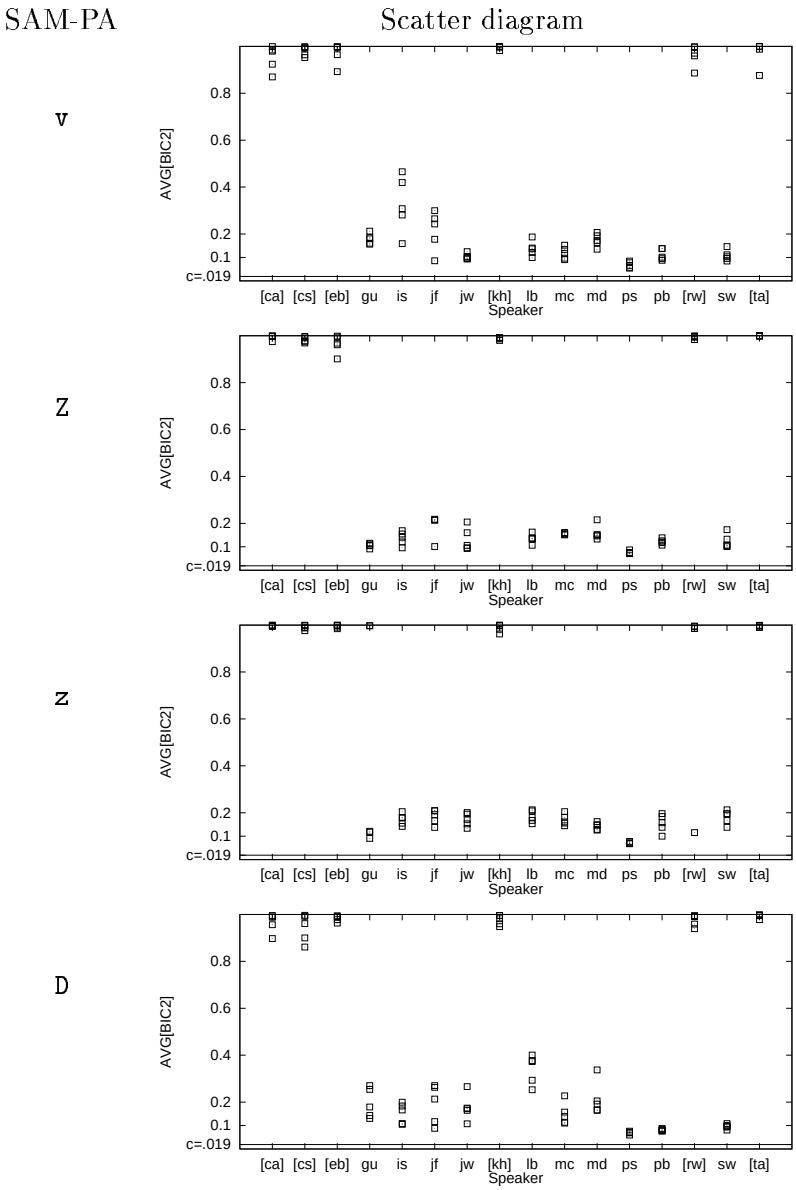

Figure 7.9: Scatter diagrams showing $\overline{b^2_{\mathrm{IT}}}$ for voiced fricatives by all 16 speakers. $f_s = 11.025$kHz. c is the critical value at the $\alpha = 0.001$ for the Gaussian hypothesis test. Female speakers are denoted by [ ].

the biphases are less significant. A hypothesis which explains this somewhat surprising result is proposed below.

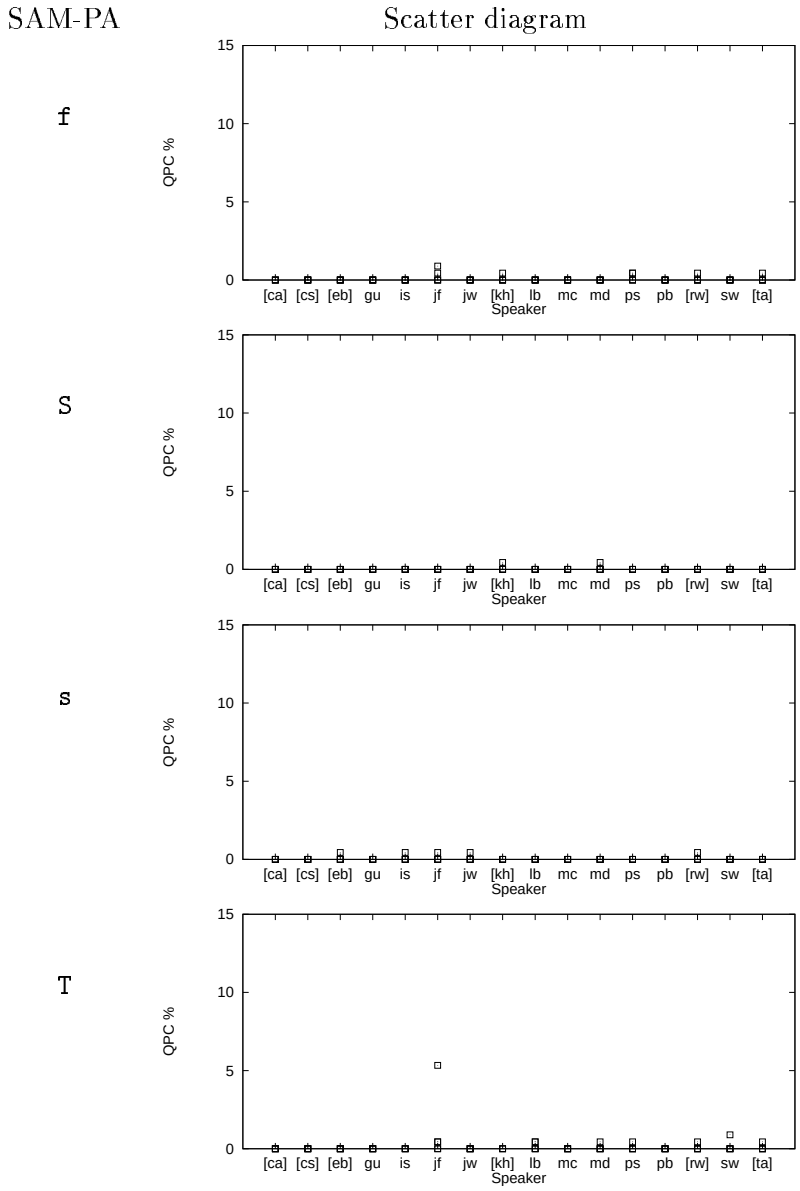

Figure 7.10: Scatter diagrams showing proportion of bifrequency bins in IT in which QPC detected for unvoiced fricatives by all 16 speakers. $f_s = 22.05$kHz. Female speakers are denoted by [ ].

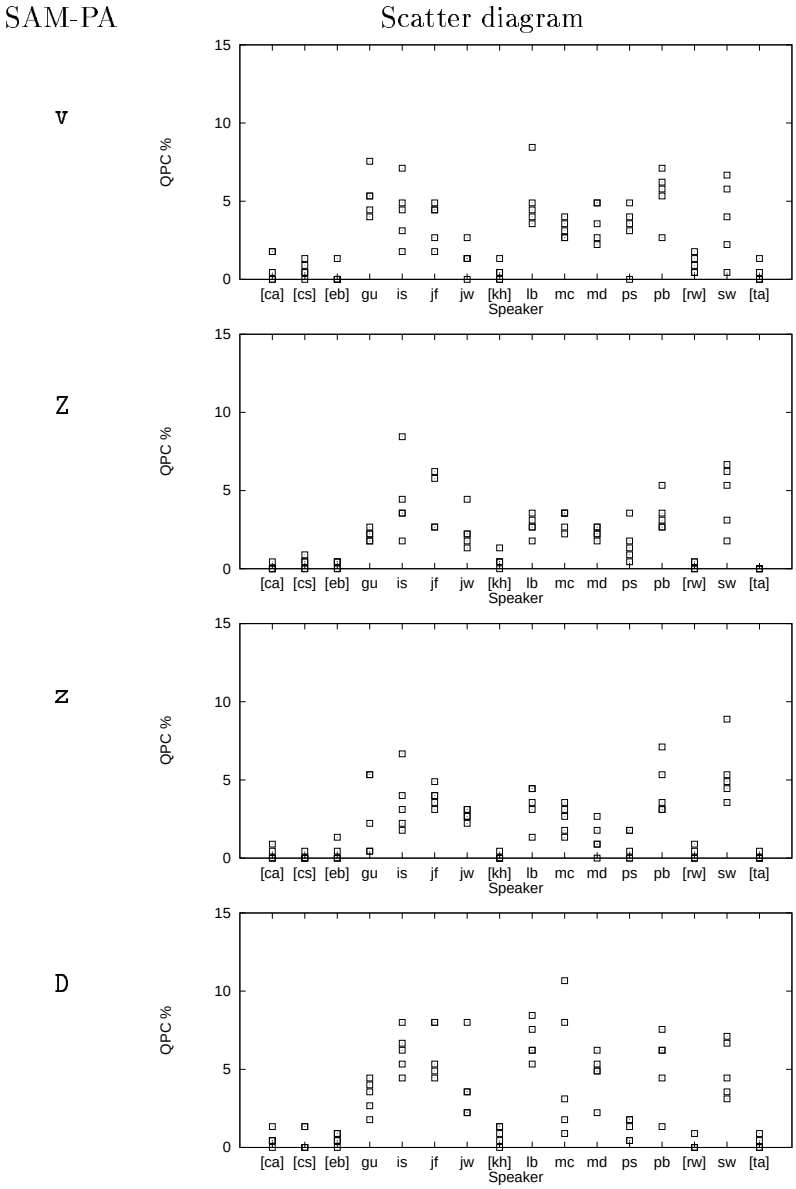

Figure 7.11: Scatter diagrams showing proportion of bifrequency bins in IT in which QPC detected for voiced fricatives by all 16 speakers. $f_s = 11.025$kHz. Female speakers are denoted by [].

7.2.3 Discussion

7.2.3.1 Unvoiced Fricatives

The average squared bicoherence $\overline{b^2_{\mathrm{IT}}}$ for the unvoiced fricatives indicates that there is a low level of quadratic phase coupling in these sounds. Interpreting the squared bicoherence as a measure of the proportion of coupled energy at each bifrequency (from Section 4.3), $\overline{b^2_{\mathrm{IT}}}$ can be interpreted as the proportion of the *total* signal energy that is quadratically coupled. For unvoiced sounds (from Figure 7.8) this comes out at about 2%, which, compared to the uncoupled proportion (the remaining 98%) corresponds to -16dB. The possibility exists that these low levels of b^2 are noise effects rather than signal attributes, but even if they are due to QPC, it seems unlikely that quadratic coupling of this magnitude can really play an important role in audible speech effects.

The threshold statistic for significant magnitude squared bicoherence has been derived from the "Gaussianity" test statistic developed in [74]. Although the single bifrequency threshold used in the analysis of *jw*'s speech was set to 0.1 (which corresponds to $\alpha = 0.001$, see Appendix A.2), the *summary* statistic for the squared bicoherence *summed* (or averaged[4]) over the IT gives a different threshold. The threshold level for the average squared bicoherence $\overline{b^2_{\mathrm{IT}}}$ with $\alpha = 0.001$ turns out to be 0.019, much lower than that for a single bifrequency bin (see Appendix A.2.3). This means that a signal having b^2 below the 0.1 threshold ($\alpha = 0.001$) at *every single* bifrequency in the IT could still have a value of $\sum_{\mathrm{IT}} b^2$ (or $\overline{b^2_{\mathrm{IT}}}$) greater than the critical threshold 0.019 ($\alpha = 0.001$), resulting in a rejection of H_0 ("Gaussianity").

Although this might appear to indicate that the squared bicoherence plots for *jw* give an under-estimate of each signal's deviation from Gaussianity, this turns out not to be the case for these speech sounds, because in Figure 7.8 it is evident that almost every unvoiced fricative for almost every speaker (except the speaker *jf*) has a value of $\overline{b^2_{\mathrm{IT}}}$ consistent with the signal being Gaussian. Even if a higher significance level is used (e.g. $\alpha = .05 \rightarrow \mathrm{Critical}[\overline{b^2_{\mathrm{IT}}}] = .017$) the majority of sounds still have $\overline{b^2_{\mathrm{IT}}}$ level below the threshold, and so H_0 is accepted for most of them. Thus from a stochastic hypothesis testing viewpoint it seems that unvoiced fricatives are Gaussian.

The conclusion to be drawn from the two ways of looking at the bicoherences of the

[4]The same arguments apply to squared bicoherence summed over the IT, or averaged over the IT, as can be seen from a consideration of the results in Appendix A.2.

unvoiced fricatives is thus that they exhibit very low levels (if any) of coupling, and that they nearly all have bicoherences consistent with the signals being Gaussian. It can thus be concluded that :

Unvoiced fricatives have no interesting bispectral properties.

Although these results have been obtained in a more rigorous fashion than previously published work [36], the conclusion reached, that the unvoiced fricatives are Gaussian, is the same. From the point of view of speech modelling, it implies that as far as unvoiced fricatives are concerned, quadratically nonlinear signal production mechanisms are unlikely to yield substantial improvements over existing techniques, and noise robust techniques based on the bispectrum are unlikely to be successful.

7.2.3.2 Voiced Fricatives

The interpretation of the bicoherence of voiced fricatives, and the other voiced sounds, is different to that of the unvoiced fricatives. For these sounds, $\overline{b^2_{\mathrm{IT}}}$ can no longer be taken as an indicator of the proportion of QPC, and can instead only be used to discriminate the bifrequencies which are dominated by deterministic components (which are candidates for exhibiting coupling) from noise processes (which are not).

The indicators from the preliminary study of *jw*'s speech are (see Figure 7.5) that voiced fricatives have very high levels of b^2, which is consistent with the signals containing a high proportion of deterministic components. However, the biphase-testing part of the QPC detector results in very few QPC detections (see Figure 7.7), and the detections that are made are scattered in an undiscernible pattern across the IT.

The analysis of the complete database yields the result that the values of $\overline{b^2_{\mathrm{IT}}}$ are very high (close to 1) for females, but around 0.2 for males (see Figure 7.9). This result occurs for all four of the voiced fricatives tested, with no apparent difference between the different sounds. However, the subsequent QPC detection test produces the surprising result (see Figure 7.7) that the females have very low numbers of QPC detections compared to the men.

This can be explained in the following way. It has been shown in Section 4.5 how the biphase detector is based on an approximate empirical expression for the variance of the biphase estimate. This led to the definition of a "acceptance region" for QPC plotted

in the complex bicoherence domain (see Figure 4.8). Referring back to this plot it can be seen that the variance of the biphase estimate is largest for lower values of b^2; if the estimate is very "clean" then b^2 will be close to unity and the width of the acceptance region will be small; if the estimate is "noisy" then b^2 will be lower and the width of the acceptance region will be large.

Now the investigation of the properties of the QPC detector (see Section 4.5.3) revealed that the probability of a Type II error (i.e. detecting QPC when it is in fact absent) is *proportional* to the width of the acceptance region. Furthermore it was shown (Figure 4.10) that as the SNR decreases, and so b^2 decreases, so the probability of making Type II errors *increases*.

There are two hypotheses to explain these results for the voiced speech. By examining the consequences of each, an explanation for the observed results will be formed. The hypotheses are :

H_{QPC} If speech signals do exhibit QPC, then it might be expected that the number of QPC detections would be determined by the number of bifrequencies which have significant b^2 magnitude. In other words, the signals which have more b^2 magnitude would be expected to have more QPC detections than the signals with low levels of b^2.

H_{UC} If the signals do not exhibit QPC, then the number of QPC detections would be determined by false alarms. From the shape of acceptance region shown in Figure 7.12 it is expected that the signals which have more b^2 magnitude will have a *lower* false-alarm rate, and so lower numbers of QPC detections.

Now the trend observed in the analysis of the voiced fricatives is that the females, who have high levels of b^2, have low levels of QPC detections, whereas the males, with low levels of b^2, have high levels of QPC detections. These findings are commensurate with the hypothesis H_{UC} above, that the speech signals exhibit no QPC. The fact that the females have higher levels of b^2 is due to the fact that female speech contains more high-frequency energy ("signal") than male speech, and this results in a situation shown schematically in Figure 7.12.

The conclusion reached then, on the basis of this explanation of the observed results is that :

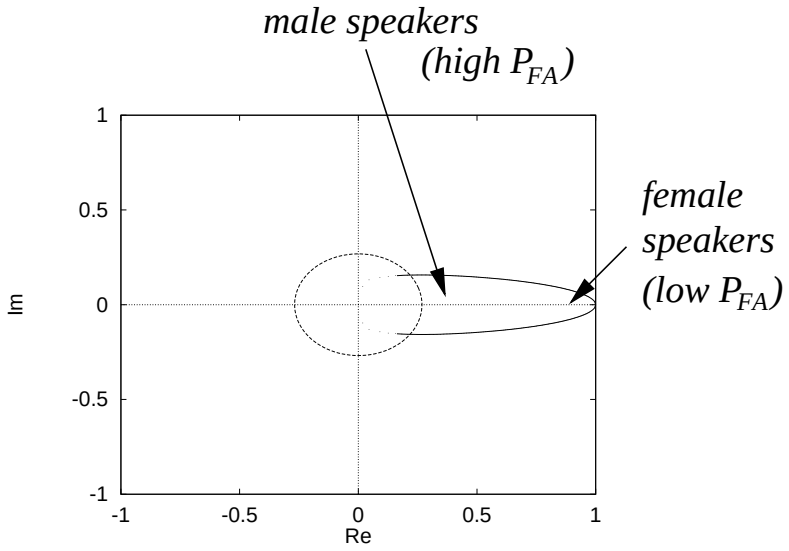

Figure 7.12: QPC acceptance region showing difference in width of acceptance region for male and female speakers.

The QPC detection rates are consistent with the hypothesis that voiced fricatives exhibit no QPC.

7.3 Vowels

7.3.1 A First Look

To get some idea of the sort of results expected from the analysis of the vowel sounds, preliminary results from a few speech sounds will be described first. For the speaker *jw*, these sounds are the vowel sounds `i`, `u`, `A`, {, corresponding approximately to the four corners of the vowel trapezium [1, p198], and `V` (schwa), corresponding to the centre of the vowel trapezium. This subset of sounds provides a fairly representative cross-section of the realisable sounds.

Figure 7.13 shows the formant chart for all 5 recorded utterances of these 5 sounds spoken by *jw*. From the formant frequencies, obtained by LPC analysis within the Entropic Waves speech analysis package, $f_2 - f_1$ is plotted against f_1. This plotting technique ensures that the formant chart can be related to the high-low front-back system used by phoneticians [1], and this labelling is also shown in the figure. The figure indicates that the consistency of the estimated formant frequencies for multiple utterances of the same sound is good.

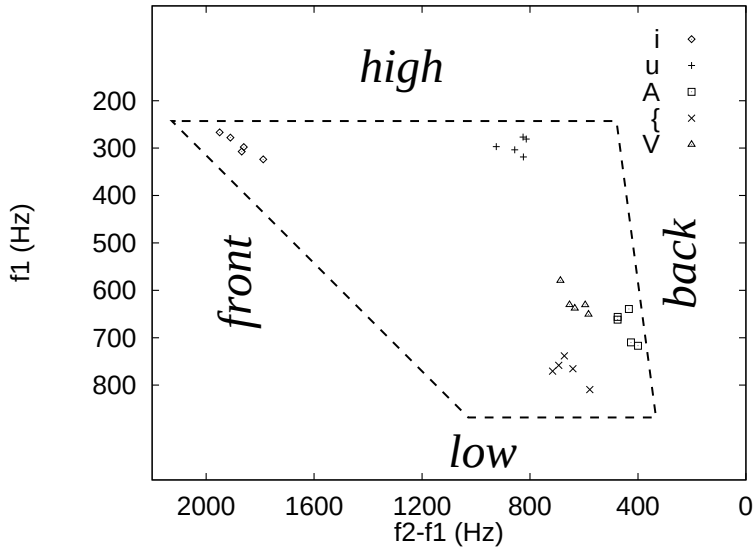

Figure 7.13: Formant chart showing 5 instances each of the vowel sounds `i`, `u`, `A`, {, and `V`, spoken by *jw*.

Figure 7.14 shows the power spectra of each of the 5 utterances of each word for this speaker. It is once again evident that the consistency between utterances is good, since power spectra of repeated utterances are very similar.

The easiest way to compare the squared bicoherences of these sounds is to look at contour plots of b^2. Once again using a threshold contour at the $\alpha = 0.001$ level, Figure 7.15 shows the squared bicoherences of these sounds. The figure shows that most of these speech samples have very high bicoherence magnitude (the $\alpha = 0.001$ contour lines at $b^2 = 0.1$ often lie right on the perimeter of the IT, indicating that squared bicoherence *everywhere* in the IT is above this level.) although there do not seem to be any discernible patterns which distinguish between the different sounds.

These results corroborate the findings of previous researchers [36][5] , that vowel sounds have high levels of squared bicoherence. However, as it was discussed in Section 4.4.1, the squared bicoherence of voiced speech sounds *is not* a reliable measure of QPC because of the issue of phase randomisation. Therefore no inference can be made as to whether these levels of b^2 indicate anything about QPC.

The QPC detections for the same sounds are shown in Figure 7.16. It is evident that even though nearly all sounds had very high levels of squared bicoherence, the number of QPC detections per sound is rather low (For example, the QPC detection plot at

[5] For the purposes of comparison, the frequency resolution in this analysis is $\Delta f = 11025/64 = 172$Hz, compared to $\Delta f = 100$Hz in [36].

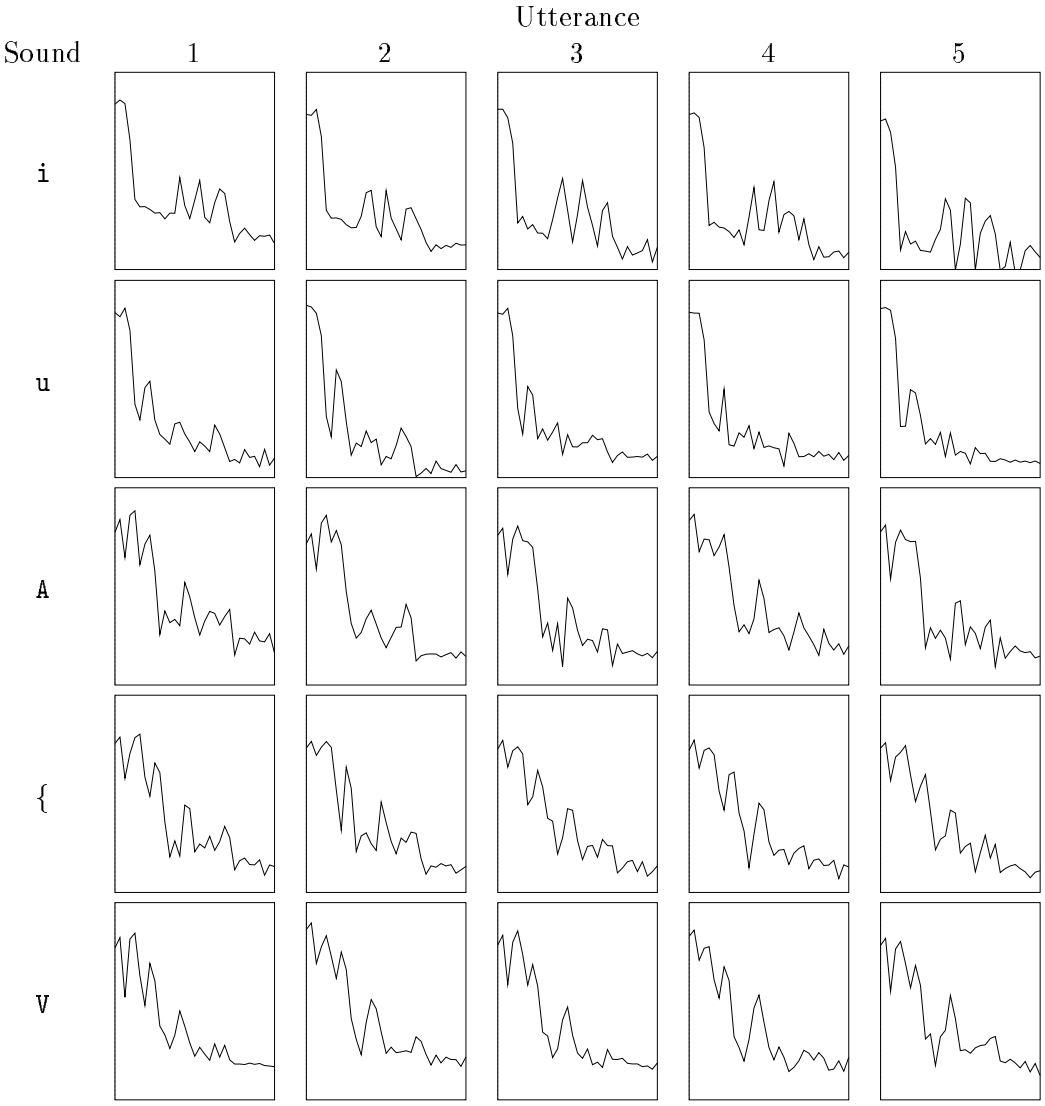

Figure 7.14: Power spectra for 5 utterances of 5 vowel sounds for speaker *jw*. For each plot, the x-axis (frequency) ranges from 0-5.5125 kHz, and the y-axis ranges from -10 to 50 dB.

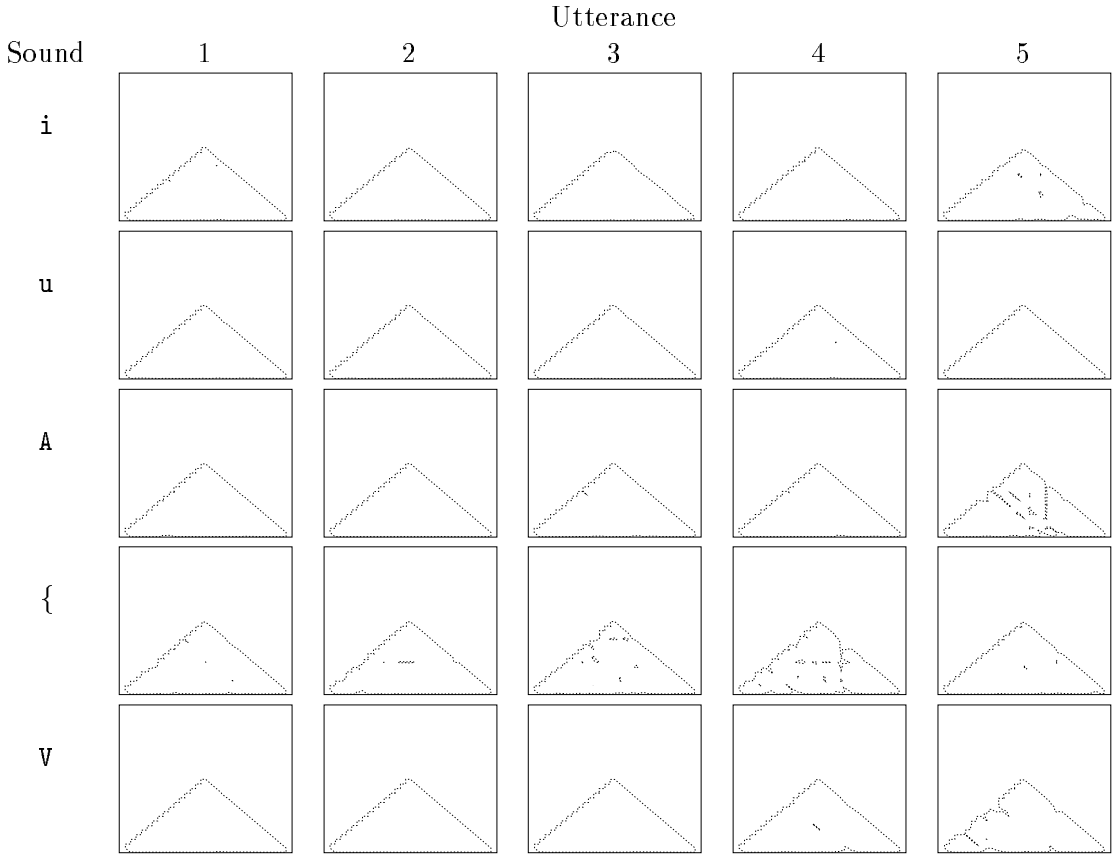

Figure 7.15: Squared bicoherences for 5 utterances of 5 vowel sounds for speaker *jw*. For each plot, the two frequency axes range from 0-5.5125 kHz ($f_s = 11.025$kHz), only the IT is shown, and the contours are at 0.1, which is the $\alpha(1) = 0.001$ significance level for squared bicoherence..

the top left corner of Figure 7.16 shows 5 detections.) And the figure also shows that there are no discernible patterns in the QPC detections. Each repeated utterance of a sound seems to result in a very different QPC detection plot.

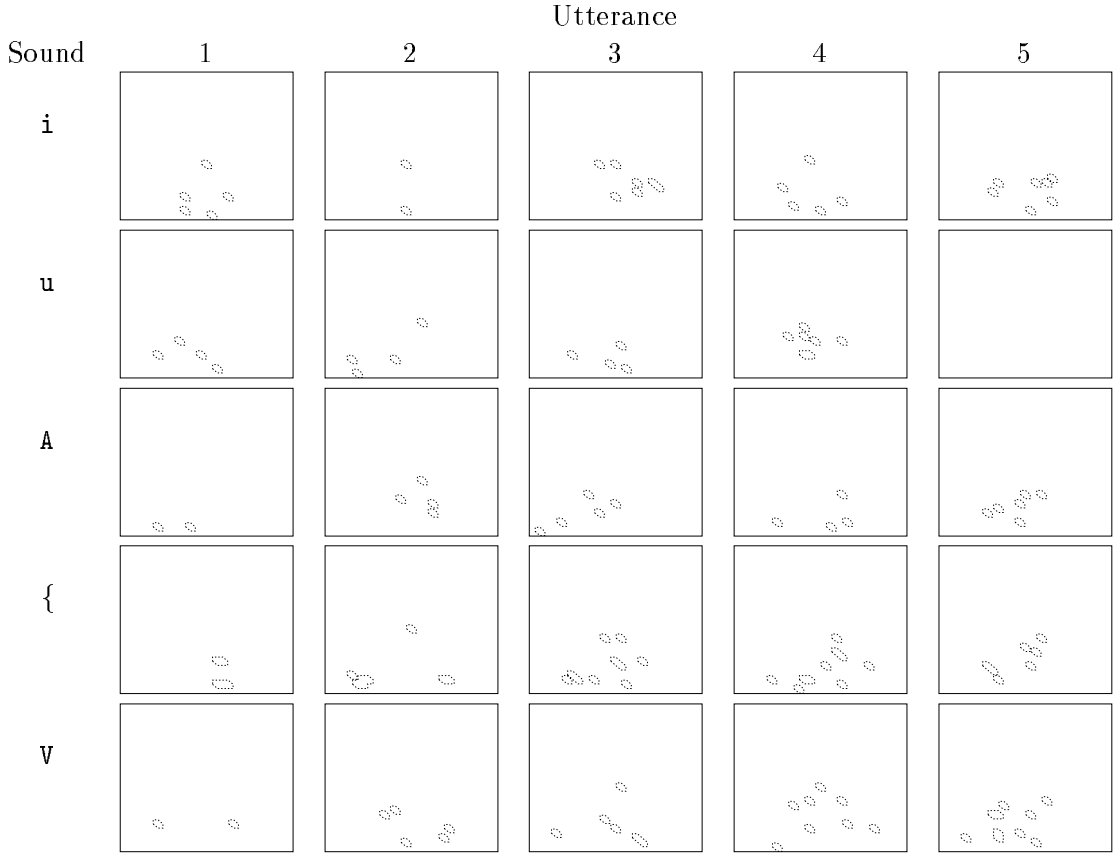

Figure 7.16: QPC detections for 5 utterances of 5 vowel sounds for speaker *jw*.

7.3.2 Descriptive Statistics

A more extensive picture can be seen by considering the descriptive statistics over the whole vowel database. The scatter diagrams for $\overline{b^2_{\mathrm{IT}}}$ (the average b^2 over the IT) are shown in Figure 7.17 for the subset of 5 vowel sounds `i`, `u`, `A`, { and `V` and the critical level $c = 0.019$ is shown on each scatter plot.

The striking difference between male and female speakers which was observed for voiced fricatives is not repeated here, but most speakers have $\overline{b^2_{\mathrm{IT}}}$ values which are significant at the 0.001 level (i.e. $\overline{b^2_{\mathrm{IT}}} > 0.019$). This indicates that a large proportion of the bifrequencies in the IT have significant b^2 levels, which means that a large proportion are tested for zero biphase in the two-part QPC test. Figure 7.18 shows the QPC

detections rates for the same vowels. It is perhaps surprising, given the high level of b^2, that the QPC detection rates are not higher - QPC detections are found in between 0 and 10% of the bifrequencies for all the vowel sounds.

7.3.3 Discussion

The pattern of results observed for the vowel sounds is not as clear as that for the voiced fricatives. There is no clear distinction between the speech of females and males, and so at this stage, no conclusion can be drawn about whether the QPC detections made are detections of actual QPC, or false alarms. However, some further analysis will be reported in Section 7.5 which throws further light on these results.

7.4 Nasals

7.4.1 A First Look

Figure 7.19 shows the power spectra of the three nasal sounds for the speaker *jw*. Once again the main comment to make from this plot is that there appears to be a fair consistency between multiple utterances of each sound. Figure 7.20 shows that, in a similar way to the voiced fricatives and vowels (in Figures 7.5 and 7.15 respectively), the nasals have very high levels of b^2, with just a few bifrequencies having b^2 values below unity. Completing the analysis, Figure 7.21 shows the QPC detections for the same sounds : once more the number of QPC detections is relatively low, with roughly less than 10 detections per sound. And again there does not appear to be a discernible pattern in these detections which distinguishes between the different nasals.

7.4.2 Descriptive Statistics

Figure 7.22 shows the scatter diagrams for $\overline{b^2_{\mathrm{IT}}}$ (b^2 averaged over the IT) for the three nasal sounds. In a similar way to the results for the voiced fricatives, the female speakers have high levels of bicoherence magnitude compared to the male speakers, although the differences between the sexes is not as great as for the voiced fricatives.

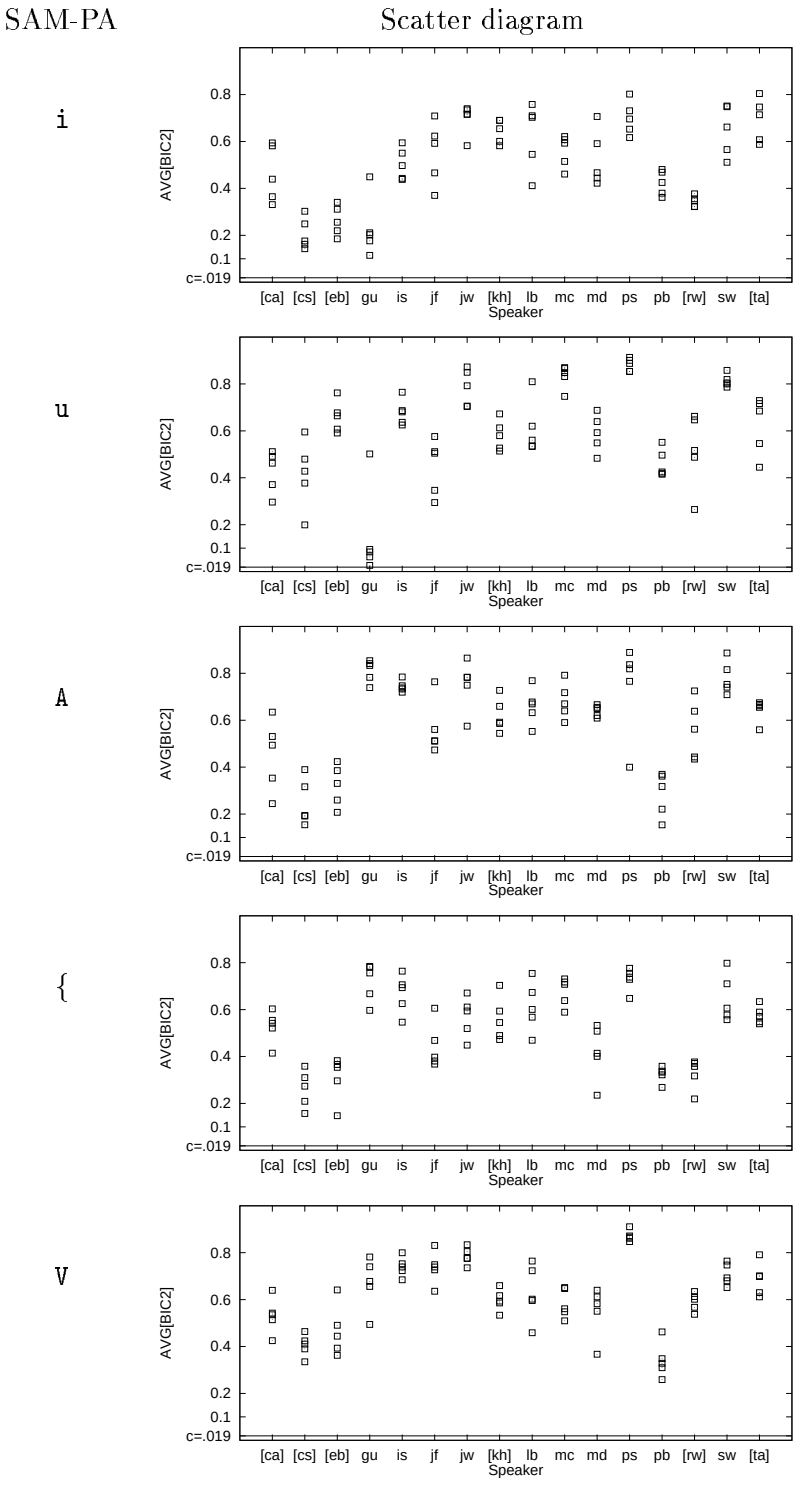

Figure 7.17: Scatter diagrams showing $\overline{b^2_{\mathrm{IT}}}$ for 5 key vowel sounds spoken by all 16 speakers. c is the critical value at the $\alpha = 0.001$ for the Gaussian hypothesis test. Female speakers are denoted by [ ].

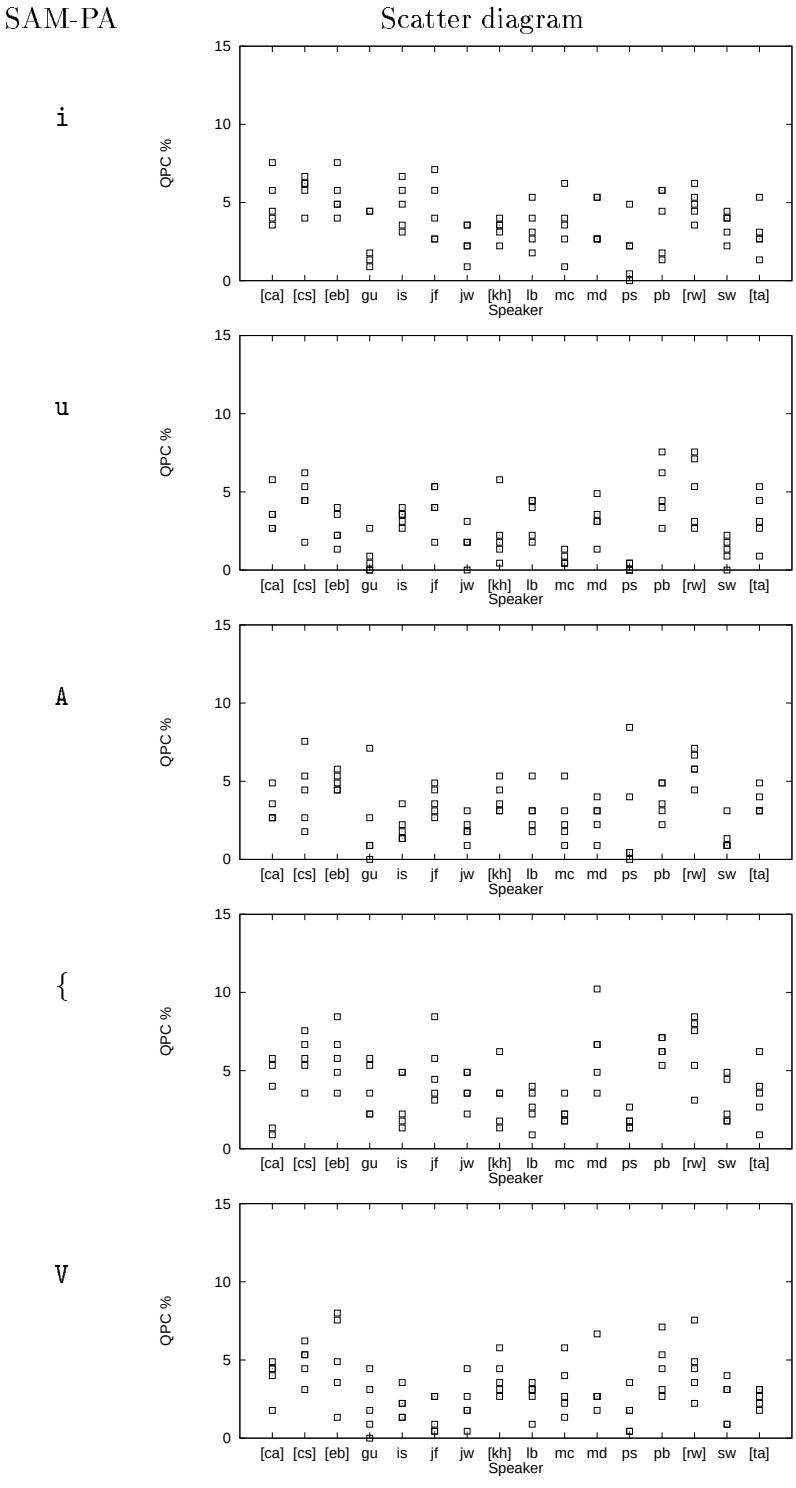

Figure 7.18: Scatter diagrams showing proportion of bifrequency bins in IT in which QPC detected for 5 key vowel sounds spoken by all 16 speakers. Female speakers are denoted by [ ].

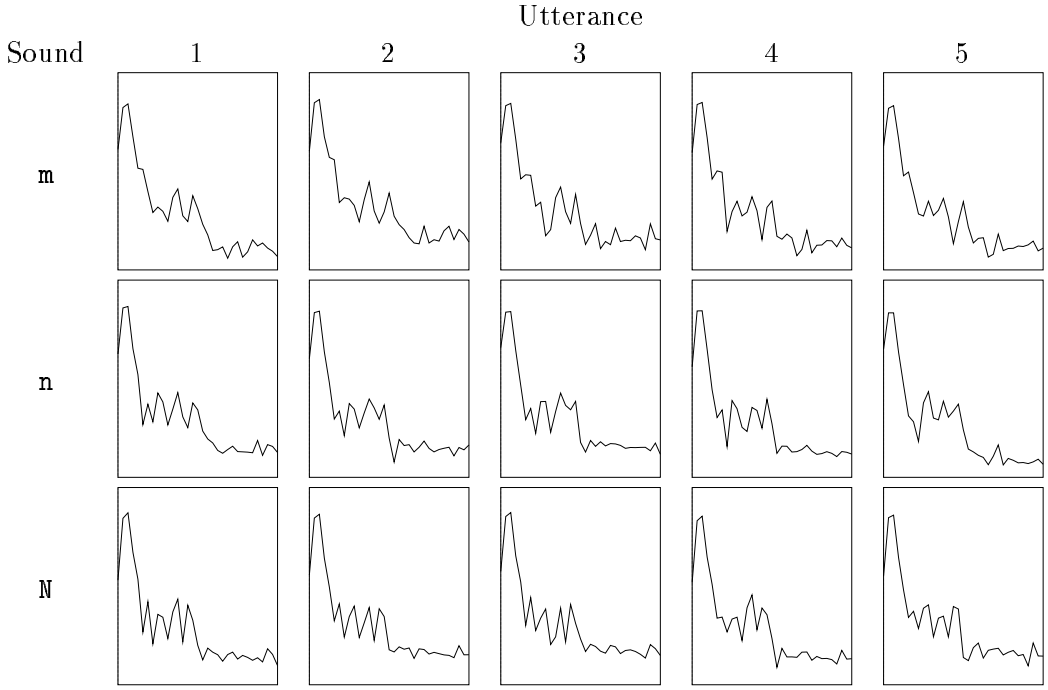

Figure 7.19: Power Spectra for 5 utterances of 3 nasal sounds for speaker *jw*. For each plot, the x-axis (frequency) ranges from 0-5.5125 kHz, and the y-axis ranges from -10 to 50 dB. $f_s = 11.025$kHz, $N = 4096$, $M = 64$, Hamming window.

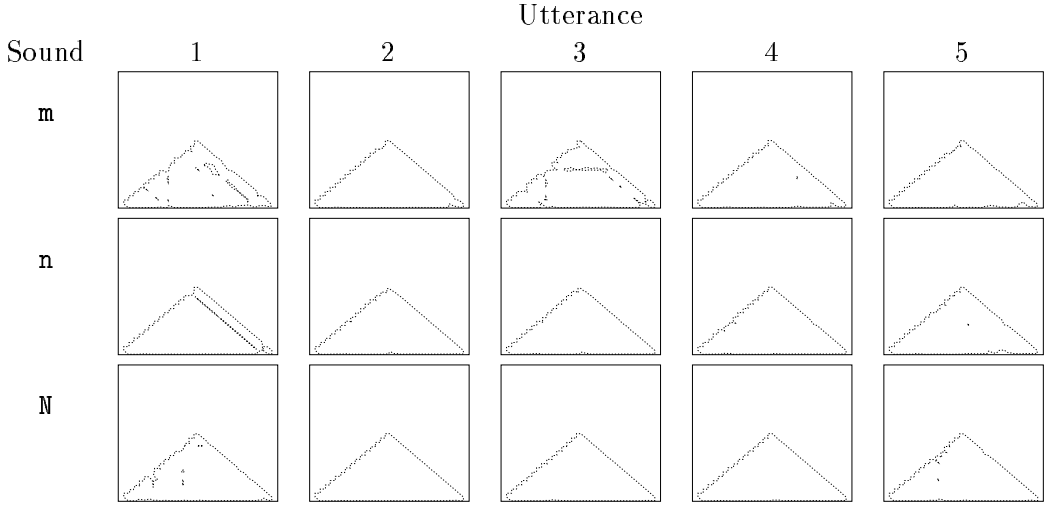

Figure 7.20: Squared bicoherences for 5 utterances of 3 nasal sounds for speaker *jw*. For each plot, the two frequency axes range from 0-5.5125 kHz, only the IT is shown, and the contours shown are at the 0.5 level.

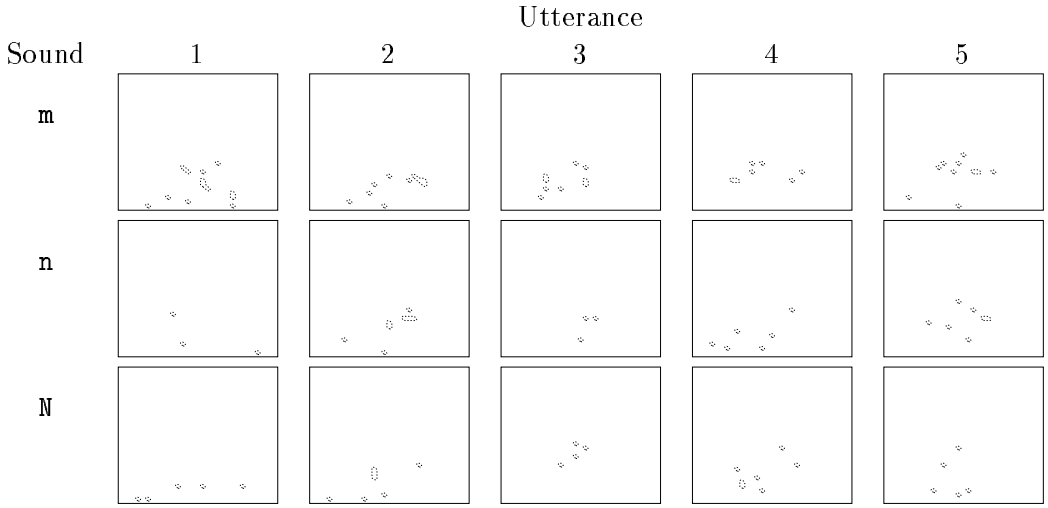

Figure 7.21: QPC detections for 5 utterances of 3 nasal sounds for speaker *jw*. For each plot, the two frequency axes range from 0-5.5125 kHz, only the IT is shown, and QPC detections are shown in black.

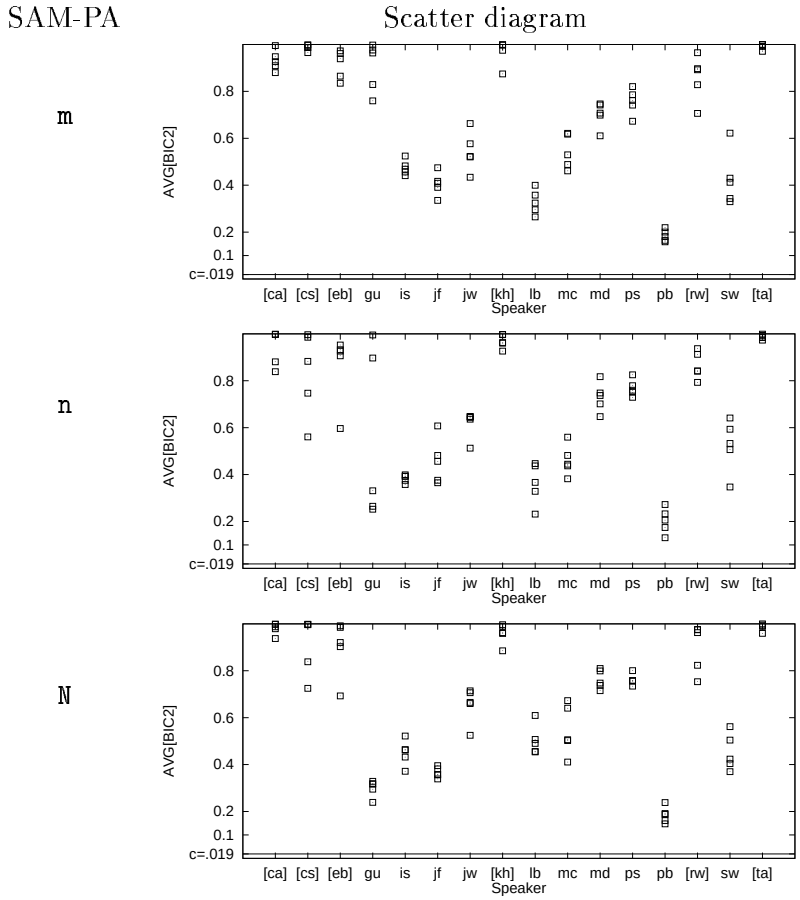

Figure 7.22: Scatter diagrams showing $\overline{b^2_{\mathrm{IT}}}$ for nasals spoken by all 16 speakers. c is the critical value at the $\alpha = 0.001$ for the Gaussian hypothesis test. Female speakers are denoted by [].

7.4.3 Discussion

It appears from these results that the level of QPC detections for the nasal sounds is quite low (with detections usually at less than 10% of the bifrequencies in the IT). These QPC detection rates reflect the pattern observed earlier for the voiced fricatives - the speakers which have high average levels of bicoherence (particularly the female speakers) have very low levels of QPC detections. The explanation for these results is thus the same as that used for the voiced fricatives - that the QPC detections which are seen are just false-alarms :

The QPC detection rates are consistent with the hypothesis that nasals exhibit no QPC.

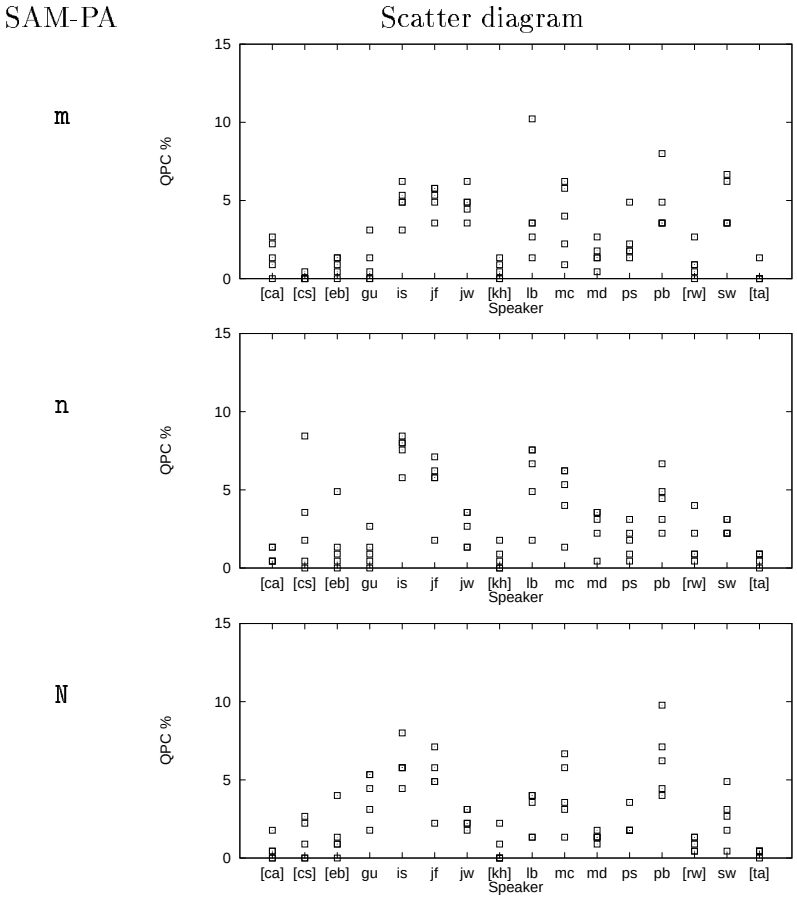

Figure 7.23: Scatter diagrams showing proportion of bifrequency bins in IT in which QPC detected for nasal sounds spoken by all 16 speakers. Female speakers are denoted by [ ].

7.5 Correlation between features

One way to find out whether the HOS-based measures used in this work are related to other conventional measures is to calculate these quantities for a number of signals and then calculate the correlation between them. This analysis has been carried out on each of the signal types - fricatives, vowels and nasals, and this section describes the results of this analysis.

In this section a number of sample correlation matrices will be shown; each entry in the correlation matrix $\mathbf{R}$ shows r_{jk}, the Pearson coefficient of correlation [107] between the two features j and k. To make these matrices easier to interpret, the rows and columns of the correlation matrices will be labelled, as shown below. The matrix is given by

$$\mathbf{R} \triangleq \begin{array}{c} \\ k=1 \\ k=2 \\ \vdots \\ k=p \end{array} \begin{array}{c} \begin{array}{cccc} j=1 & j=2 & \dots & j=p \end{array} \\ \begin{bmatrix} r_{11} & & & \\ r_{21} & r_{22} & & \\ \vdots & \vdots & \ddots & \\ r_{p1} & r_{p2} & \dots & r_{pp} \end{bmatrix} \end{array} \tag{7.1}$$

where

$$\begin{aligned} r_{jk} &\triangleq \frac{c_{jk}}{\sqrt{c_{jj}c_{kk}}} \\ c_{jk} &\triangleq \sum_{i=1}^{n}(x_{ij}-\overline{x}_j)(x_{ik}-\overline{x}_k)/(n-1), \end{aligned} \tag{7.2}$$

in which n is the number of samples (i.e. utterances available) and $x_{i1}, x_{i2}, .., x_{ip}$ are the p features which are measured for the ith utterance.

Each coefficient ranges between -1 (strong negative correlation) and +1 (strong positive correlation). Under the (rather strong) assumption that each pair of variables are together normally distributed, the significance of these coefficients can be measured. In the matrices which follow, coefficients significant at the 0.05 level are shown in **bold**, and coefficients not significant at the 0.05 level are shown in *italic*.

7.5.1 Correlation between b^2 and s^2

It was mentioned in Chapter 3 that the squared bicoherence and skewness function, two of the available bispectrum normalisation schemes, are very similar to one another. To verify this, it is interesting to look at the correlation between the squared bicoherence and the skewness function in the IT. Also of interest, although it has not been discussed in this thesis to any great depth, is the relation between the squared bicoherence (in the IT) and the skewness function in the OT[6].

Table 7.1 shows the correlation matrix between $\sum_{\mathrm{IT}} b^2$: the squared bicoherence summed in the IT, $\sum_{\mathrm{IT}} s^2$, $\sum_{\mathrm{OT}} s^2$: the skewness function summed in the IT and OT respectively, and $\sum_{\mathrm{IT}} q$: the number of QPC detections in the IT (using the detector described in Chapter 4), for all the unvoiced and voiced fricatives[7].

$$
\begin{array}{c}
\\
\sum_{\mathrm{IT}} b^2 \\
\sum_{\mathrm{IT}} s^2 \\
\sum_{\mathrm{OT}} s^2 \\
\sum_{\mathrm{IT}} q
\end{array}
\begin{array}{c}
\begin{array}{cccc}
\sum_{\mathrm{IT}} b^2 & \sum_{\mathrm{IT}} s^2 & \sum_{\mathrm{OT}} s^2 & \sum_{\mathrm{IT}} q
\end{array} \\
\left[\begin{array}{cccc}
\mathbf{1} & & & \\
\mathbf{.88} & \mathbf{1} & & \\
\mathbf{.64} & \mathbf{.54} & \mathbf{1} & \\
\mathbf{.63} & \mathbf{.77} & \mathbf{.39} & \mathbf{1}
\end{array}\right]
\end{array}
$$

$$
\begin{array}{c}
\\
\sum_{\mathrm{IT}} b^2 \\
\sum_{\mathrm{IT}} s^2 \\
\sum_{\mathrm{OT}} s^2 \\
\sum_{\mathrm{IT}} q
\end{array}
\begin{array}{c}
\begin{array}{cccc}
\sum_{\mathrm{IT}} b^2 & \sum_{\mathrm{IT}} s^2 & \sum_{\mathrm{OT}} s^2 & \sum_{\mathrm{IT}} q
\end{array} \\
\left[\begin{array}{cccc}
\mathbf{1} & & & \\
\mathbf{1.00} & \mathbf{1} & & \\
\mathbf{1.00} & \mathbf{1.00} & \mathbf{1} & \\
\mathbf{-.67} & \mathbf{-.67} & \mathbf{-.68} & \mathbf{1}
\end{array}\right]
\end{array}
$$

Table 7.1: Correlation coefficients between features for unvoiced (top) and voiced (bottom) fricatives.

For both voiced and unvoiced fricatives, it is clear that the squared bicoherence and skewness function (both in the IT and OT) are highly correlated, although the correlation is higher for the voiced sounds than the unvoiced ones. It can also be seen that especially for the voiced sounds, the skewness function results in the IT and OT are highly correlated.

Reinforcing the results seen in Section 7.2.2, for the voiced fricatives the QPC detections

[6] For further information on the use and interpretation of the OT see [67, 108].

[7] A total of 320 utterances (5 utterances $\times$ 4 words $\times$ 16 speakers) were used to calculate each correlation coefficient.

are *negatively* correlated with the squared bicoherence levels, i.e. the voiced fricatives with the higher bicoherences have the lower QPC detections.

Table 7.2 provides further evidence of this difference in the correlation patterns between unvoiced and voiced fricatives for HOS-based features from different IT zones. Only part of each correlation matrix is shown, illustrating how $\overline{b_i^2}(i = 1, .., 4)$ is related to $\overline{q_i}(i = 1, .., 4)$. The boxes in the correlation matrices indicate the coefficients which relate features derived from the same IT zone, since it is expected that these will be most highly correlated. The pattern is again confirmed, that in each IT zone, $\overline{b_i^2}$ and $\overline{q_i}$ are positively correlated for unvoiced fricatives and negatively correlated for voiced fricatives.

$$
\begin{array}{c} \\ \overline{q_1} \\ \overline{q_2} \\ \overline{q_3} \\ \overline{q_4} \end{array}
\begin{array}{c}
\begin{array}{cccc} \overline{b_1^2} & \overline{b_2^2} & \overline{b_3^2} & \overline{b_4^2} \end{array} \\
\left[\begin{array}{cccc}
\boxed{.69} & .47 & .26 & .20 \\
.49 & \boxed{.54} & .29 & .18 \\
.14 & .13 & \boxed{.29} & \textit{-.01} \\
\textit{.05} & \textit{.05} & \textit{.06} & \boxed{.26}
\end{array} \right]
\end{array}
$$

$$
\begin{array}{c} \\ \overline{q_1} \\ \overline{q_2} \\ \overline{q_3} \\ \overline{q_4} \end{array}
\begin{array}{c}
\begin{array}{cccc} \overline{b_1^2} & \overline{b_2^2} & \overline{b_3^2} & \overline{b_4^2} \end{array} \\
\left[\begin{array}{cccc}
\boxed{-.64} & -.67 & -.66 & -.67 \\
-.45 & \boxed{-.52} & -.54 & -.54 \\
-.47 & -.46 & \boxed{-.48} & -.48 \\
-.38 & -.38 & -.39 & \boxed{-.39}
\end{array} \right]
\end{array}
\tag{7.3}
$$

Table 7.2: Part of the correlation coefficients between subzone features for unvoiced (top) and voiced (bottom) fricatives. Coefficients not significant at the 0.05 level are in *italic*.

The correlations between the summary features $\sum_{\text{IT}} b^2$, $\sum_{\text{IT}} s^2$, *skwsumot* and *qsumit* for the other voiced sounds (the vowels and nasals as shown in Table 7.3) show similar trends to the voiced fricatives in Table 7.1. Again there is very high correlation between the squared bicoherence and skewness functions, but negative correlation between the squared bicoherence level and the number of QPC detections. For each of the three classes of voiced sounds the correlation coefficient between $\sum_{\text{IT}} b^2$ and $\sum_{\text{IT}} q$ is between -.59 and -.74. This confirms the results reported from the scatter plots above, since speakers which have high levels of $\sum_{\text{IT}} b^2$ (or $\overline{b_{\text{IT}}^2}$) have low levels of $\sum_{\text{IT}} q$.

Although in the analysis of the vowel sounds (Section 7.3.2) no specific patterns could

be distinguished between males and females, the correlation results indicate that $\sum_{\mathrm{IT}} q$ and $\sum_{\mathrm{IT}} b^2$ are strongly negatively correlated (in the same way as they are correlated for the other voiced sounds - the voiced fricatives and the nasals). Therefore the same explanation for the QPC detection rates can be adopted, namely that :

The QPC detection rates are consistent with the hypothesis that vowels exhibit no QPC.

$$
\begin{array}{c}
 \\
\sum_{\mathrm{IT}} b^2 \\
\sum_{\mathrm{IT}} s^2 \\
\sum_{\mathrm{OT}} s^2 \\
\sum_{\mathrm{IT}} q
\end{array}
\begin{array}{c}
\begin{array}{cccc}
\sum_{\mathrm{IT}} b^2 & \sum_{\mathrm{IT}} s^2 & \sum_{\mathrm{OT}} s^2 & \sum_{\mathrm{IT}} q
\end{array} \\
\left[\begin{array}{cccc}
\mathbf{1} & & & \\
\mathbf{1.00} & \mathbf{1} & & \\
\mathbf{.86} & \mathbf{.87} & \mathbf{1} & \\
\mathbf{-.59} & \mathbf{-.59} & \mathbf{-.58} & \mathbf{1}
\end{array}\right]
\end{array}
$$

$$
\begin{array}{c}
 \\
\sum_{\mathrm{IT}} b^2 \\
\sum_{\mathrm{IT}} s^2 \\
\sum_{\mathrm{OT}} s^2 \\
\sum_{\mathrm{IT}} q
\end{array}
\begin{array}{c}
\begin{array}{cccc}
\sum_{\mathrm{IT}} b^2 & \sum_{\mathrm{IT}} s^2 & \sum_{\mathrm{OT}} s^2 & \sum_{\mathrm{IT}} q
\end{array} \\
\left[\begin{array}{cccc}
\mathbf{1} & & & \\
\mathbf{.99} & \mathbf{1} & & \\
\mathbf{.93} & \mathbf{.96} & \mathbf{1} & \\
\mathbf{-.74} & \mathbf{-.73} & \mathbf{-.71} & \mathbf{1}
\end{array}\right]
\end{array}
$$

Table 7.3: Correlation coefficients between different features for vowels (top) and nasals (bottom).

7.5.2 Correlation between HOS measures and other measures

In this section some correlations between HOS measures and some conventional speech characterisation features are investigated.

7.5.2.1 Fricatives

Table 7.4 shows the correlation matrices between a variety of features for the fricative sounds (unvoiced and voiced). These features include $\sum_{\mathrm{IT}} b^2$, the four subzone averages $\overline{b_i^2}$ $(i = 1, .., 4)$, as well as the four spectral moments, all as described in Section 6.5.

Looking at the correlation matrix for the unvoiced fricatives first, it is evident that the HOS-based features are strongly correlated with each other, which is entirely as expected since the sum of the squared bicoherence over all the subzones is the same as the sum over the IT. The correlation appears to be strongest for the low frequency

subzone $\overline{b_1^2}$, and weakest for the high frequency subzone $\overline{b_4^2}$. This is probably due to the fact that the unvoiced speech sounds generally contain more energy at low frequencies (from Figure 7.3), and so the SNR in subzone 1 will be higher than that in subzone 4. The spectral moments $m_1 - m_4$ are quite strongly correlated with each other, which in itself is an interesting result, since it suggests that there is some redundancy in the measures. In particular m_1 and m_3 are strongly negatively correlated ($r = -.83$), and m_3 and m_4 are quite strongly positively correlated ($r = .68$). However, the correlation coefficients between the spectral moments and the HOS-based features are generally very low, and at the .05 level the only significant correlation value is $r = -.12$ between $\overline{b_2^2}$ and m_2.

Turning now to the correlation matrix for the voiced fricatives, a similar pattern is observed. The HOS-based features are all very highly correlated with each other (more highly than for the unvoiced fricatives). This time the correlation patterns amongst the spectral moments are different: there is still strong negative correlation between m_3 and m_1 ($r = -.69$), and strong positive correlation between m_3 and m_4 ($r = .87$), but now there is also correlation between m_2 and m_3 ($r = -.71$) and between m_1 and m_2 ($r = .93$). Finally for the voiced fricatives m_1 is weakly correlated with the HOS-based measures $\sum_{\mathrm{IT}} b^2$, $\overline{b_2^2}$, $\overline{b_3^2}$ and , $\overline{b_4^2}$.

7.5.2.2 Vowels

The correlation matrix relating HOS-based features and formant frequencies for the vowel sounds is shown in Table 7.5. Once again there is strong correlation between each of the HOS-based features, although the correlations between these features for the vowel sounds are generally slightly weaker than for the voiced fricatives described above. There are several significant correlations between the conventional measures (in this case the fundamental voicing frequency f_0 and the formant frequencies f_1 and f_2). Considering the first column of the top correlation matrix first, there is significant negative correlation between $\sum_{\mathrm{IT}} b^2$ and f_0 and strong negative correlation between $\sum_{\mathrm{IT}} b^2$ and f_2, and between $\sum_{\mathrm{IT}} b^2$ and $f_2 - f_1$, indicating that vowels sounds with high second formants tend to have low $\sum_{\mathrm{IT}} b^2$. Now referring back to Figure 7.13 and the vowel trapezium interpretion in terms of vowel "height" and "frontness"[8] it is clear that $f_2 - f_1$ is a measure of vowel *height*, and f_1 a measure of vowel *front-ness*. Hence it can be inferred that front vowels (such as `i`) have a lower $\sum_{\mathrm{IT}} b^2$ than back vowels,

[8] "height" and "frontness" refer to the position of the tongue in the VT [1].

	$\sum_{\mathrm{IT}} b^2$	$\overline{b_1^2}$	$\overline{b_2^2}$	$\overline{b_3^2}$	$\overline{b_4^2}$	m_1	m_2	m_3	m_4
$\sum_{\mathrm{IT}} b^2$	**1**								
$\overline{b_1^2}$	**.79**	**1**							
$\overline{b_2^2}$	**.77**	**.51**	**1**						
$\overline{b_3^2}$	**.68**	**.36**	**.39**	**1**					
$\overline{b_4^2}$	**.64**	**.27**	**.34**	**.32**	**1**				
m_1	*−.05*	*−.10*	*−.04*	*.09*	*−.06*	**1**			
m_2	*−.01*	*.09*	**−.12**	*−.04*	*.01*	**−.16**	**1**		
m_3	*.02*	*.05*	*.06*	*−.09*	*.02*	**−.83**	**−.12**	**1**	
m_4	*.01*	*−.01*	*.02*	*−.03*	*.03*	**−.29**	**−.38**	**.68**	**1**

	$\sum_{\mathrm{IT}} b^2$	$\overline{b_1^2}$	$\overline{b_2^2}$	$\overline{b_3^2}$	$\overline{b_4^2}$	m_1	m_2	m_3	m_4
$\sum_{\mathrm{IT}} b^2$	**1**								
$\overline{b_1^2}$	**.97**	**1**							
$\overline{b_2^2}$	**1.00**	**.96**	**1**						
$\overline{b_3^2}$	**1.00**	**.95**	**1.00**	**1**					
$\overline{b_4^2}$	**1.00**	**.95**	**.99**	**1.00**	**1**				
m_1	**.11**	*.04*	**.12**	**.13**	**.13**	**1**			
m_2	*.03*	*.04*	*.03*	*.05*	*.05*	**.93**	**1**		
m_3	*−.01*	*.02*	*.00*	*−.02*	*−.01*	**−.69**	**−.71**	**1**	
m_4	*.04*	*.04*	*.05*	*.03*	*.04*	**.48**	**.51**	**.87**	**1**

(7.4)

Table 7.4: Correlation coefficients between features for unvoiced (top) and voiced (bottom) fricatives. Coefficients significant at the 0.05 level shown in **bold**, coefficients not significant at the 0.05 level shown in *italic*.

and also that the vowel *height* (measured by f_1) does not appear to have much effect on the level of $\sum_{\mathrm{IT}} b^2$. The $\sum_{\mathrm{IT}} q$ results (lower correlation matrix in Table 7.5) reflect the patterns already observed of negative correlation between squared bicoherence and QPC detections, and so this matrix does not really provide any new useful information.

7.5.3 Discussion

No particularly interesting correlations exist between the HOS-based measures developed in this thesis and the conventional measures such as spectral moments (for fricatives) and formant frequencies (for vowels). This could be taken as an indication that the HOS-based measures are working on totally separate signal properties (e.g. coupling) to the conventional measures (spectral shape), but perhaps a more reasonable explanation is that, since no significant evidence of coupling (or non-Gaussian be-

	$\sum_{\mathrm{IT}} b^2$	$\overline{b_1^2}$	$\overline{b_2^2}$	$\overline{b_3^2}$	$\overline{b_4^2}$	f_0	f_1	f_2	$f_2 - f_1$
$\sum_{\mathrm{IT}} b^2$	**1**								
$\overline{b_1^2}$	**.81**	**1**							
$\overline{b_2^2}$	**.94**	**.82**	**1**						
$\overline{b_3^2}$	**.97**	**.71**	**.90**	**1**					
$\overline{b_4^2}$	**.91**	**.57**	**.77**	.84	**1**				
f_0	**−.25**	*.03*	**−.13**	**−.21**	**−.43**	**1**			
f_1	*−.05*	*.07*	*−.04*	*−.02*	**−.12**	**.28**	**1**		
f_2	**−.40**	**−.24**	**−.33**	**−.39**	−.44	**.21**	**−.26**	**1**	
$f_2 - f_1$	**−.35**	**−.23**	**−.28**	**−.34**	**−.37**	**.11**	**−.50**	**.97**	**1**

	$\sum_{\mathrm{IT}} q$	$\overline{q_1}$	$\overline{q_2}$	$\overline{q_3}$	$\overline{q_4}$	f_0	f_1	f_2	$f_2 - f_1$
$\sum_{\mathrm{IT}} q$	**1**								
$\overline{q_1}$	**.44**	**1**							
$\overline{q_2}$	**.65**	**.17**	**1**						
$\overline{q_3}$	**.74**	**.13**	**.32**	**1**					
$\overline{q_4}$	**.75**	**.12**	**.24**	**.40**	**1**				
f_0	**.24**	*−.08*	**.09**	**.18**	**.34**	**1**			
f_1	**.14**	*−.02*	*.06*	*.08*	**.19**	**.28**	**1**		
f_2	**.34**	**.10**	**.23**	**.23**	**.29**	**.21**	**−.26**	**1**	
$f_2 - f_1$	**.27**	**.09**	**.19**	**.18**	**.21**	**.11**	**−.50**	**.97**	**1**

(7.5)

Table 7.5: Correlation coefficients between features for vowels. Coefficients significant at the 0.05 level shown in **bold**, coefficients not significant at the 0.05 level shown in *italic*. The top matrix shows correlations between features based on b^2, the bottom matrix shows correlations between features based on q.

haviour) has been found in the speech sounds, then the bispectral measures of such signals carry minimal information.

7.6 Robust Estimator

All the analyses described above have also been carried out using the robust bicoherence estimator described in Chapter 5. Perhaps because of the facts that the speech data was recorded very carefully under studio conditions, and that some noisy files were excluded from analysis, the results of bicoherence estimation with the robust detector are very similar to the results obtained using the conventional estimator, since the outlier detection algorithm detects very few outliers, and so for these signals the outlier removal part of the algorithm remains mostly dormant. The fact that this happens is in fact one of the good properties of the outlier rejection algorithm - the data is only

manipulated **if** outliers are detected.

Figures 7.24 and 7.25 show the scatter plot, by speaker, of the values of average robust b^2 and QPC detections for the subset of five vowels - these figures are the robustly estimated analogues of Figures 7.17 and 7.18, and a comparison between these two pairs of plots shows that, for these sounds, the robust estimator has very little effect.

The scatter plots in Figures 7.24 and 7.25 can be compared to the their non-robust equivalents in Figures 7.17 and 7.18 (in Section 7.3.2), and are described in Section 7.6.

The fact that the robust estimator has little effect on the estiamted HOS measures is not too discouraging - the recording operation was carried out very carefully, and so the dangers of transient contamination were minimised. The robust techniques may prove to be more useful in fields in which transient contamination is more of a problem, such as underwater acoustics.

7.7 Further Analysis

More sophisticated analyses such as Analysis of Variance (ANOVA) are valid only if the data has certain properties which allow relevant assumptions to be made. One of these, which is a prerequisite of ANOVA, is that the variances of the quantities of interest are homogeneous across speakers. The validity of this assumption can be tested using the Levene test[109, 110]. In Appendix D.1 the results of applying the Levene test to the feature $\sum_{\mathrm{IT}} b^2$ are described. The results indicate that for fricatives, vowels, and nasals, the hypothesis of homogeneity of variance is *rejected* for most of the data. This means that the variances of the summed squared bicoherence for different speakers are different, and rules out ANOVA analysis unless the data is transformed in some way. Whether a suitable transformation could be used is beyond the scope of the current investigation, but could be a topic for further investigation.

A final comment is necessary regarding the pitches of utterances in the database. As it was mentioned in Appendix C.1.1, the voicing pitch was not controlled during the database collection. It is easy to argue that if the voicing pitch changes during an utterance then each new pitch cycle will occur in a different acoustic environment to the preceding one. This raises the possibility of non-stationarity accross the utterance, a property with which the segment-averaging estimation approach used here cannot cope. However, although this problem has not been catered for explicitly in the experimental

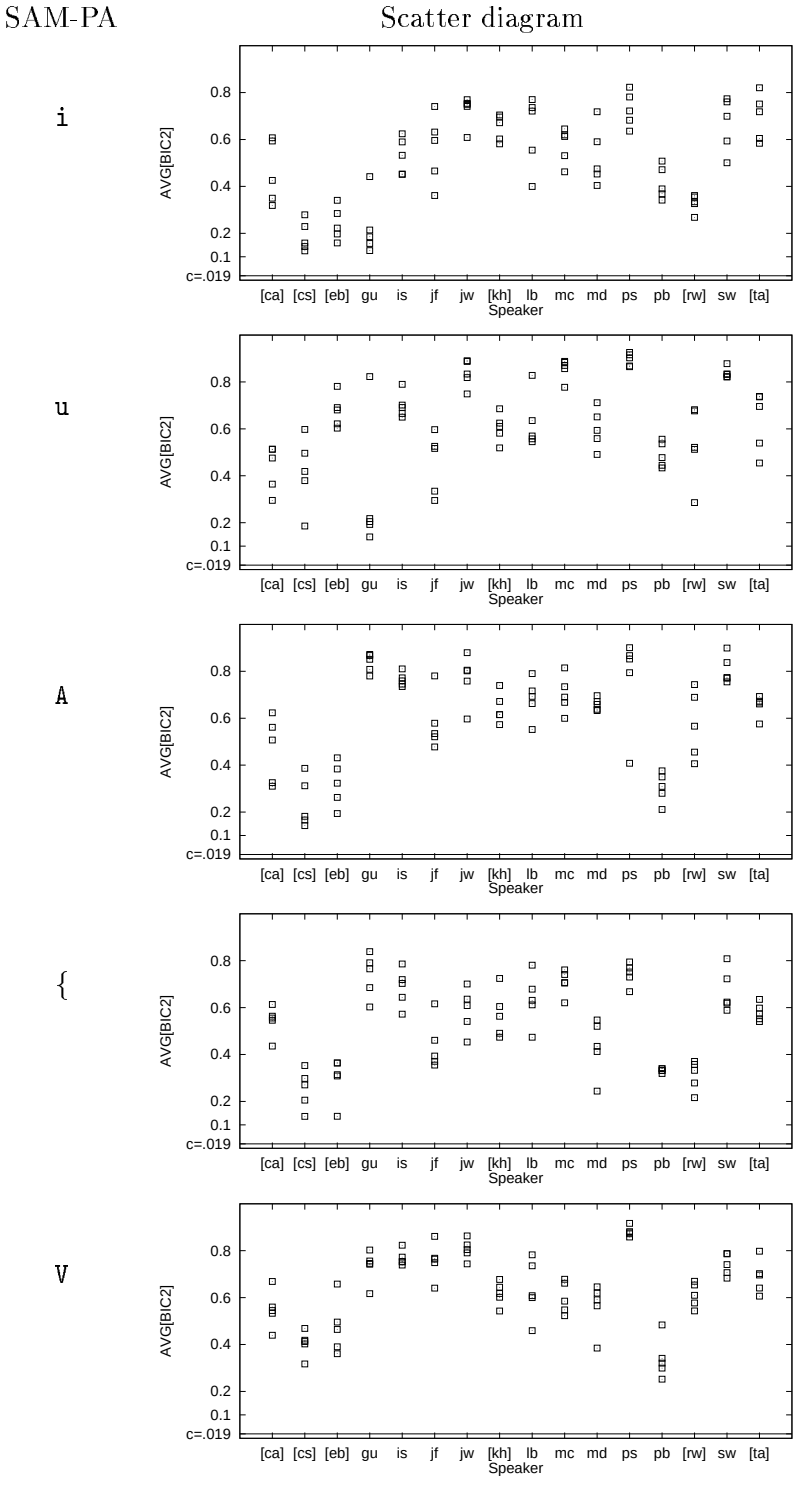

Figure 7.24: Scatter diagrams showing robust $\overline{b^2_{\mathrm{IT}}}$ for 5 key vowel sounds spoken by all 16 speakers. c is the critical value at the $\alpha = 0.001$ for the Gaussian hypothesis test. Female speakers are denoted by [ ].

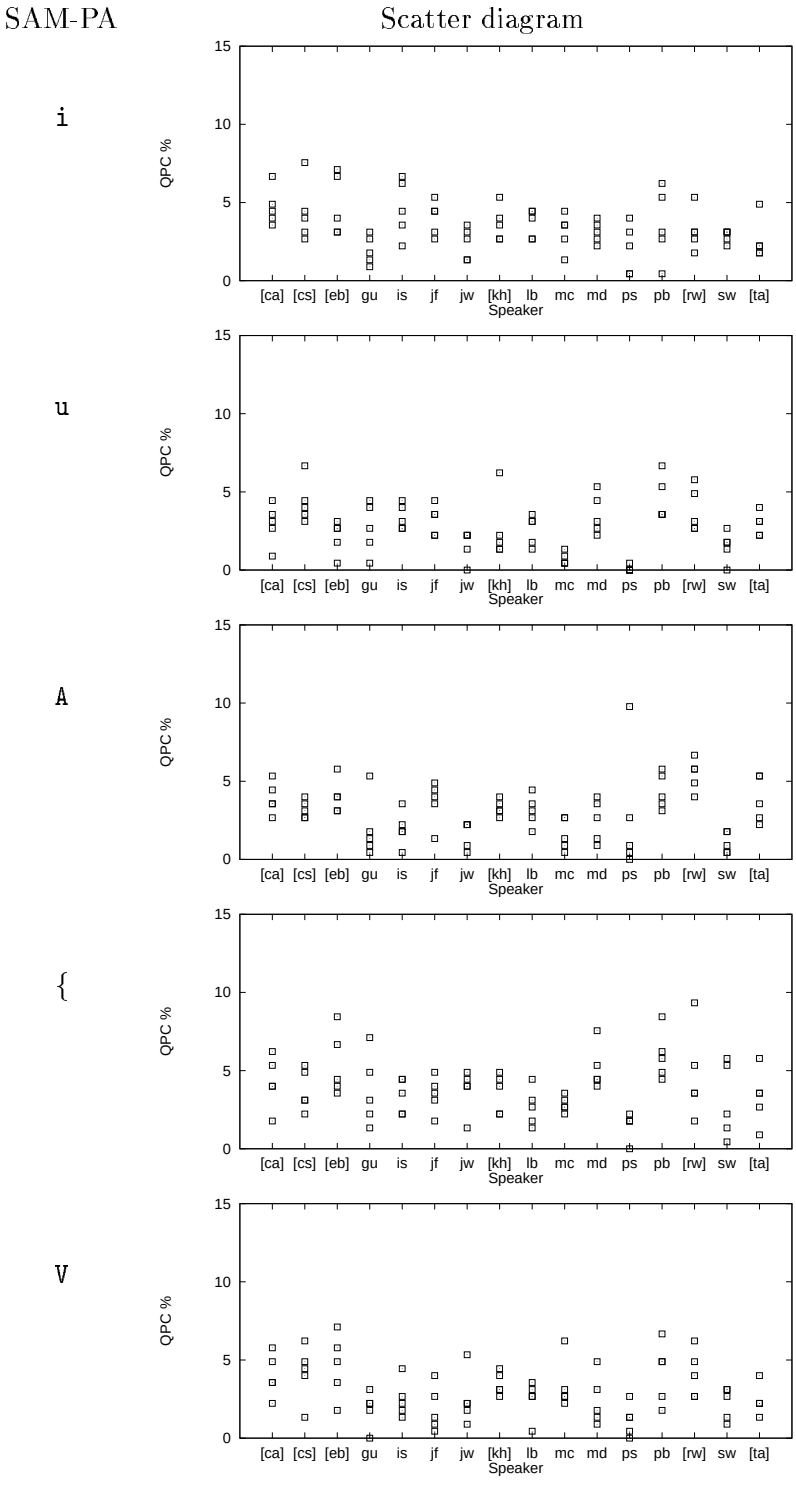

Figure 7.25: Scatter diagrams showing proportion of bifrequency bins in IT in which robust QPC detected for 5 key vowel sounds spoken by all 16 speakers. Female speakers are denoted by [ ].

design, it is unlikely that it has caused any problems in the analysis, because the vast majority of utterances in the database do have approximately flat pitch. The investigation of how pitch changes affect the HOS measures could, however, be an interesting area for further work.

7.8 Summary

This chapter has described the results of applying the bispectral tools developed in earlier chapters to the specially constructed speech database. The results indicate that unvoiced fricatives have bispectrum properties consistent with them being Gaussian signals (i.e. their bicoherence levels are not statistically different from zero), but that most voiced sounds have bicoherences much larger than would be expected under the Gaussian hypothesis. However, when the two-part QPC detector developed in Chapter 4 is applied to the voiced speech data, the number of detections of QPC in the IT is very small, and it is seen that the speakers who have the highest levels of squared bicoherence also have the lowest levels of QPC. In confirmation of this, negative correlation has been observed between the magnitude of the squared bicoherence and the number of QPC detections. A hypothesis to explain this phenomenon has been proposed - that this number of QPC detections is consistent with the signals not exhibiting QPC, i.e. that the detections which do occur are false alarms (Type II errors). Thus the conclusion is reached that the speech sounds do not exhibit significant levels of QPC.

Chapter 8

Conclusion

At the beginning of this thesis the strategy defined was firstly to attempt to identify and characterise quadratic nonlinearities in speech signals, and secondly, to see how this information could be used in future speech processing systems. The results indicate that there do not appear to be any quadratic nonlinearities in the speech sounds measured, so the second step has not been taken. This approach has resulted in a thesis with a somewhat negative conclusion, but it is believed that this is preferable to the approach, sometimes taken in this field, of attempting to utilise nonlinear properties which speech signals may not have.

In the pursuit of some meaningful bispectral analysis of speech, this thesis has described new developments in several areas of higher order statistics. Some conclusions will now be drawn about the main achievements of this work, its limitations, and suggestions for areas in which further work is required.

In Chapter 3 an attempt has been made to pull together the previous work on nonparametric bispectrum estimation, to highlight the similarities, and a few of the differences, between the different measures which have been proposed. By presenting the properties of normalised bispectra (skewness and squared bicoherence) in a unified way, this should provide a guide to the somewhat conflicting use of terminology in the literature. The discussion of estimation issues is often neglected in theoretical developments, so the considerations of the effects of data length and data windows (the first extension of results concerning windowing effects to normalised bispectra) is also important.

Chapters 4 and 5 contain the main theoretical development in this thesis. Chapter 4 outlines problems which exist for conventional bispectrum estimators which can render them ambiguous as measures of QPC. In the light of this discovery, some results in published literature must be reappraised. It is certainly likely that some of the author's own previously-published work overestimated the extent of QPC in machine signals. To

remedy this problem a novel visualisation of the problem has been presented, which suggests detecting QPC using the biphase instead of the squared bicoherence magnitude. In order to reduce the number of false detections, the proposed detector consists of two parts; the first removes from future consideration noisy bifrequencies, whilst the second detects biphase. This detector has been developed using empirical expressions for the statistical properties of the biphase, rather than theoretical expressions. Although this represents a weakness in the work as far as mathematical rigour is concerned, it has been demonstrated that, for certain signals, the expressions which define this detector can be rewritten in a form which is very close to the rigorous asymptotic treatment provided recently by Zhou et al [4, 5].

The detector based on the asymptotic mathematics does however, differ from the one described here in one critical aspect : Zhou's detector uses a single record of data, with no averaging, but the detector described here uses extensive averaging in order to achieve reliable estimates. Common sense dictates that for finite data length records, averaged estimates are less susceptible to measurement noise, even if the asymptotic mathematics suggest that this is not the case, and further work is needed to extend the mathematical framework of the single-shot detector to encompass the segment-averaged estimate.

A further contribution, which can be applied to either of the detectors described above, concerns the exposition of the problem of Type II errors in QPC detection. A new expression has been presented for the probability of false alarm which can be used for either of the two detectors described. This expression suggests that the performance of biphase-based QPC detectors degrades as the SNR decreases, and that this degradation is manifested by an increasing rate of Type II errors. If the SNR becomes too low the approximation on which the biphase-based detector is based breaks down, and so the performance of the detector based on this approximation falls rapidly, because false detections occur so often. Although the explanation of the mechanism for this degradation in performance is very simple, it does seem to be reasonably accurate, since the trends predicted by this theory have been observed in simulation experiments.

The work reported in Chapter 5 was prompted by observations of instability in the bispectra of some real-life signals. This is believed to be the first attempt at using robust estimation techniques for bispectrum estimation, and the results described are very encouraging. Through a series of thought experiments a new transient-resistant estimator has been developed which has attractive properties; it allows for correlation

between the real and imaginary parts of the bispectral estimate; it maintains the useful bounded property of the squared bicoherence; and it appears, from simulation results, to work very effectively. The algorithm requires no prior knowledge about the location or variance of the transients, but it does require that they do not occur too often. This is not too serious a limitation, because if an experimental signal contains very many transients, then the question arises that maybe these transients are actually part of the "signal" and not the "noise".

In retrospect, further work on the robust detector is unlikely to be justified in the speech field, since in the results described in Chapter 7 the robust estimator gave very similar results to the ordinary estimator, this result being attributable to the high quality of the recorded speech. However, other signal processing fields, such as underwater acoustics and machine signature analysis, may prove to be more fruitful for the robust estimator.

Chapter 6 describes preliminary work carried out prior to the main speech analysis, and this includes a new investigation into some of the issues surrounding the application of bispectral analysis in a pitch-synchronous framework, the description of a special speech database designed for reliable bispectral estimation, as well as covering experimental considerations such as filter properties for reliable bispectral estimation. It is hoped that this work may be useful to others wishing to carry out similar types of speech analysis.

Chapter 7 represents the first comprehensive analysis of speech sounds using the bispectrum, from a nonlinearity detection perspective. The approach taken in all this work has been to carefully develop understanding of bispectral phenomena, and so the speech analysis can be appraised in the light of this new understanding.

The findings for unvoiced fricatives indicate that these sounds do not have significant bicoherences, and so the use of third-order HOS measures for noise-robust processing of such sounds is unlikely to prove to be useful. The voiced sounds do have significant bicoherence magnitudes, but from the development in Chapter 4 this is now understood to be a sufficient, but not a necessary, condition for QPC. Indeed, from the considerations of the origins of Type II errors in biphase detection from Chapter 4, the biphase detection results are fully consistent with speech not exhibiting any quadratic coupling. This would appear to rule out quadratic nonlinearities for speech production modelling.

As a result of this work, the bispectral estimation process is now better understood.

The underlying problem which hinders the use of bispectral methods is the fact that they require such long data records, and this in turn increases the chance that non-stationarities will arise in the signal which distort the estimates. This is not a new problem, nor is it an easily solved one. One of the lessons that the signal processing community has learned in its dealings with HOS techniques is how demanding these techniques are in terms of data record lengths. Given that HOS analysis is so fraught with difficulties, it makes no sense to use HOS techniques unless the signals of interest (or the noise of interest) have some interesting HOS properties. The evidence from this investigation is that continuant speech signals do not have any interesting HOS properties at third-order, and therefore that there are unlikely to be any advantages in implementing speech processing algorithms in the third-order HOS domain.

It is however, relevant to note here that even though there is nothing interesting in these speech sounds at third-order, it does not follow that there is nothing interesting at fourth-order. Indeed fourth-order properties may still contain useful information about deviations from Gaussian pdfs and cubic phase coupling which nonlinear speech production models can use. The problem is that the fourth-order measures are (even) more difficult to reliably estimate than the third-order measures, and this may in fact rule out such approaches for speech signals because of the rapidly changing nature of speech.

As a further caveat, the fact that no interesting third-order properties have been found in voiced and unvoiced continuants does not rule out the utility of third-order HOS techniques for other classes of speech sounds. But once again the problem of estimation looms, since the bispectra of non-continuant sounds are more difficult to estimate that those of continuant sounds, again because of the data lengths involved.

The practice of applying pre-emphasis to speech signals, which was avoided in this work, needs to be examined from a HOS perspective, to see if any real improvements in estimates can be obtained. Other issues also need to be addressed ; the issue of devising reliable HOS estimates from short data records, very useful for speech analysis, has not been touched on in the current work, neither has the the issue of the effects of coarticulation on the HOS measures. The pattern which emerges again and again is that the well-established techniques which are used for second-order analysis of speech sounds need to be re-examined with great care when the analysis is based on HOS.

The question of whether extra speech naturalness can be achieved by using nonlinear models for speech remains an open one, and there is room for a great deal more re-

search into this interesting area. However, the work described in this thesis appears to indicate that nonlinear models based on quadratic nonlinearites do not appear to be good candidates for the job.

Finally, it is stressed that this thesis is not all about speech processing. Chapters 3- 5 are in many ways independent of speech processing frameworks, and so the techniques developed therein can be applied to other types of signals. In fact speech processing is probably one of the most difficult fields in which to carry out HOS estimation. In other fields, such as machine condition monitoring, the experimenter has much more control over the data record length, and so the problems of nonstationarity are reduced. It is felt that because of their poor estimation properties, HOS measures may be better suited to fields such as these, and that development work of HOS tools may progress in a more steady fashion in these areas.

References

[1] P. Ladefoged, *A Course in Phonetics.* Harcourt Brace Jovanovich, San Diego, 1975.

[2] J. N. Holmes, *Speech synthesis and recognition.* Van Nostrand Reinhold, Berkshire, UK, 1988.

[3] M. B. Priestley, *Spectral Analysis and Time Series.* London: Academic Press, Harcourt Brace Jovanovich, 1992.

[4] G. Zhou, G. B. Giannakis, and A. Swami, "HOS for processes with mixed spectra," in *IEEE Signal Processing ATHOS Workshop on Higher-Order Statistics*, (Begur, Girona, Spain), pp. 352–356, IEEE, June 1995.

[5] G. Zhou and G. B. Giannakis, "Polyspectral analysis of mixed processes and coupled harmonics," *IEEE Transactions on Information Theory*, vol. 42, pp. 943–958, May 1996.

[6] J. W. A. Fackrell, S. McLaughlin, and P. R. White, "Practical issues in the application of the bicoherence for the detection of quadratic phase coupling," in *IEEE Signal Processing ATHOS Workshop on Higher-Order Statistics*, (Begur, Girona, Spain), pp. 310–314, June 1995.

[7] K. K. Paliwal and M. M. Sondhi, "Recognition of noisy speech using cumulant based linear prediction analysis," in *International Conference on Acoustics, Speech and Signal Processing*, (Toronto, Canada), pp. 429–432, 1991.

[8] J. R. Deller, J. G. Proakis, and J. H. L. Hansen, *Discrete-time processing of speech signals.* New York: Macmillan Publishing Company, 1993.

[9] J. H. Eggen, *On the quality of synthetic speech :evaluation and improvements.* PhD thesis, Technical University of Eindhoven, The Netherlands, September 1992. Obtained from the AUTHOR June 1994.

[10] R. Linggard, *Electronic synthesis of speech.* Cambridge University Press, 1985.

[11] S. R. Paget, *Human Speech : Some Observations, Experiments and Conclusions as to the Nature, Origin, Purpose and Possible Improvement of Human Speech.* London: Kegan Paul, Trench, Trubner and Co Ltd, 1930.

[12] A. Breen, "Speech synthesis models : a review," *Electronics and Communication Engineering Journal*, pp. 19–31, February 1992.

[13] L. C. Wood and D. J. B. Pearce, "Excitation synchronous formant analysis," *IEE Proc I*, vol. 136, pp. 110–118, April 1989.

[14] H. L. F. Helmholtz, *On the sensations of tone as a physiological basis for the theory of music.* New York: Dover, 1954. translation of 4th German edition 1877.

[15] P. M. Milner, *Physiological Psychology.* New York Holt Rinehart and Winston, 1970.

[16] H. E. White and D. H. White, *Physics and Music : The Science of Musical Sound.* Philadelphia: Saunders College, 1980.

[17] R. Plomp and H. J. M. Steeneken, "Effect of phase on the timbre of complex tones," *Journal of the Acoustical Society of America*, vol. 46, no. 2, pp. 409–421, 1969.

[18] N. S. Jayant and P. Noll, *Digital Coding of Waveforms.* New Jersey: Prentice-Hall Inc, 1984.

[19] G. Gabor and Z. Gyorfi, "On the higher order distributions of speech signals," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, no. 4, pp. 602–603, 1988.

[20] A. Falaschi and I. Tidei, "Speech innovation characterization by higher order moments," in *Visual Representations of Speech Signals* (M. Cooke, S. Beet, and M. Crawford, eds.), pp. 243–250, John Wiley, 1993.

[21] G. Jacovitti and A. F. P Pierucci, "Speech segmentation and classification using higher order moments," in *Eurospeech '91*, pp. 1335–1338, 1991.

[22] H. M. Teager, "Some observations on oral air flow during phonation," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-28, pp. 599–601, October 1980.

[23] H. M. Teager and S. M. Teager, "Evidence for nonlinear sound production mechanisms in the vocal tract," in *Speech Production and Speech Modelling* (W. J. Hardcastle and A. Marchal, eds.), pp. 241–261, Kluwer Academic Publishers, 1990.

[24] G. Kubin, "Nonlinear processing of speech," in *Speech Coding and Synthesis* (W. B. Kleijn and K. K. Paliwal, eds.), pp. 557–610, New York: Elsevier Science, 1995.

[25] L. Atlas and J. Fang, "Quadratic detectors for general nonlinear analysis of speech," in *International Conference on Acoustics, Speech and Signal Processing*, vol. 2, (San Francisco, USA), pp. 9–11, IEEE, 1992.

[26] M. Banbrook, *Nonlinear dynamics analysis of speech from a synthesis perspective.* PhD thesis, University of Edinburgh, Edinburgh, UK, 1996. (submitted July 1996).

[27] B. Townshend, "Nonlinear prediction of speech," in *International Conference on Acoustics, Speech and Signal Processing*, (Toronto, Canada), pp. 425–429, IEEE, 1991.

[28] J. Thyssen, *Non-linear analysis, prediction, and coding of speech.* PhD thesis, Tele Danmark Research, Hørsholm, Denmark, July 1995.

[29] J. Thyssen, H. Nielsen, and S. D. Hansen, "Non-linearities in speech," in *Proceedings of the 1995 IEEE Workshop on Nonlinear Signal and Image Processing* (I. Pitas, ed.), vol. 2, (Neos Marmaras, Halkidiki, Greece), pp. 662–665, June 1995.

[30] M. D. Godfrey, "An exploratory study of the bispectrum of economic time series," *Journal of the Royal Statistical Society*, vol. 14, no. Series C Applied Statistics, pp. 48–69, 1965.

[31] R. A. Haubrich, "Earth noise, 5 to 500 millicycles per second, 1. spectral stationarity, normality and nonlinearity," *Journal of Geophysical Research*, vol. 70, pp. 1415–1427, March 1965.

[32] P. L. Brockett, M. J. Hinich, and G. R. Wilson, "Non-linear and non-Gaussian ocean noise," *Journal of the Acoustical Society of America*, vol. 82, pp. 1386–1394, October 1987.

[33] T. Sato, K. Sasaki, and Y. Nakamura, "Realtime bispectral analysis of gear noise ad its application to contactless diagnosis," *Journal of the Acoustical Society of America*, vol. 62, pp. 382–387, August 1977.

[34] Y. C. Kim and E. J. Powers, "Digital bispectral analysis and its applications to nonlinear wave interactions," *IEEE Transactions on Plasma Science*, vol. PS-7, no. 2, pp. 120–131, 1979.

[35] A. W. Lohmann and B. Wirnitzer, "Triple correlations," *Proceedings of the IEEE*, vol. 72, pp. 889–901, July 1984.

[36] B. B. Wells, "Voiced/unvoiced decision based on the bispectrum," in *International Conference on Acoustics, Speech and Signal Processing*, pp. 1589–1592, 1985.

[37] J. M. Mendel, "Tutorial on Higher Order Statistics (Spectra) in signal processing and system theory: theoretical results and some applications," *Proceedings of the IEEE*, vol. 79, pp. 278–305, March 1991.

[38] J. Vidal, E. Masgrau, A. Moreno, and J. A. R. Fonollosa, "Speech analysis using higher order statistics," in *Visual representations of speech signals* (M. Cooke, S. Beet, and M. Crawford, eds.), pp. 347–354, John Wiley, 1993.

[39] A. Moreno, J. A. R. Fonollosa, and J. Vidal, "Vocoder design based on HOS," in *Eurospeech '93*, (Berlin, Germany), pp. 519–522, September 1993.

[40] J. M. Salavedra, E. Masgrau, A. Moreno, and X. Jove, "A speech enhancement system using higher order AR estimation in real environments," in *Eurospeech '93*, (Berlin, Germany), pp. 223–226, ESCA, September 1993.

[41] J. M. Salavedra, E. Masgrau, A. Moreno, and X. Jove, "Comparison of different order cumulants in a speech enhancement system by adaptive wiener filtering," in *IEEE Signal Processing Workshop on Higher-Order Statistics*, (Lake Tahoe, California, USA), pp. 61–65, IEEE, 1993.

[42] R. H. Wang and Y. F. Liu, "Speech analysis based on bispectrum," in *Proc Int Conference on Signal Processing, Beijing, China*, pp. 373–376, 1990.

[43] W.-T. Chen and C.-Y. Chi, "Deconvolution and vocal tract parameter estimation of speech signals by higher order statistics based inverse filters," in *IEEE Signal Processing Workshop on Higher-Order Statistics*, (Lake Tahoe, California, USA), IEEE, June 1993.

[44] B. Boyanov, S. Hadjitodorov, and T. Ivanov, "Analysis of voiced speech by means of bispectrum," *Electronics Letters*, vol. 27, no. 24, 1991.

[45] B. Boianov, "Analysis of pathological voice," Tech. Rep. EEC Research Contract ERB-CIPA-CT-92-0170, Ecole Nationale Superieure des Telecommunications, 1993.

[46] S. Dubnov, N. Tishby, and D. Cohen, "Bispectrum of musical sounds : an auditory perspective," in *Proceedings of the X Colloquium on Musical Informatics*, (Milan, Italy), 1993. obtained via FTP July 1995.

[47] I. J. Gdoura, P. Louzou, and A. Spanias, "Speech processing using higher order statistics," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, vol. 1, (Chicago), pp. 160–163, May 1993.

[48] R. Fulchiero and A. S. Spanias, "Speech enhancement using the bispectrum," in *International Conference on Acoustics, Speech and Signal Processing*, vol. IV, (Minneapolis, USA), pp. 488–491, 1993.

[49] M. J. Hinich and E. Shichor, "Bispectral analysis of speech," in *Proceedings of the 17th Convention of electrical and electronic engineers in Israel*, pp. 357–360, 1991.

[50] S. Seetharaman and M. E. Jernigan, "Speech signal reconstruction based on higher order spectra," in *International Conference on Acoustics, Speech and Signal Processing*, pp. 703–706, 1988.

[51] M. Rangoussi, A. Delopoulos, and M. Tsatsanis, "On the use of Higher-Order Statistics for robust endpoint detection of speech," in *IEEE Signal Processing Workshop on Higher-Order Statistics*, (Lake Tahoe, California, USA), pp. 56–60, IEEE, June 1993.

[52] M. Rangoussi, S. Bakamidis, and G. Carayannis, "Robust endpoint detection of speech in the presence of noise," in *Eurospeech '93*, (Berlin, Germany), pp. 649–652, ESCA, September 1993.

[53] A. Moreno and J. A. R. Fonollosa, "Cumulant-based voicing decision in noise corrupted speech," in *International Conference on Spoken Language Processing*, pp. 531–534, October 1992.

[54] J. L. Navarro, A. Moreno, and E. LLeida, "Bispectral-based statistics applied to speech endpoint detection," in *IEEE Signal Processing ATHOS Workshop on Higher-Order Statistics*, (Begur, Girona, Spain), pp. 280–283, June 1995.

[55] J. L. Navarro-Mesa and A. Moreno, "Skewness and nonstationary measures applied to reliable speech endpoint detection," in *Eurospeech '95*, (Madrid, Spain), pp. 1423–1426, ESCA, September 1995.

[56] M. C. Dogan and J. M. Mendel, "Real time robust pitch detector," in *International Conference on Acoustics, Speech and Signal Processing*, (San Francisco, USA), pp. I129–I132, 1992.

[57] A. Moreno and J. A. R. Fonollosa, "Pitch determination of noisy speech using HOS," in *International Conference on Acoustics, Speech and Signal Processing*, (San Francisco, USA), pp. 133–136, 1992.

[58] Y. Kamp and C. Ma, "Connection between weighted LPC and HOS for AR model estimation," in *Eurospeech '93*, (Berlin, Germany), pp. 345–347, ESCA, September 1993.

[59] D. R. Brillinger, "Some basic aspects and uses of higher-order spectra," *Signal Processing (Eurasip)*, vol. 36, pp. 239–249, 1994.

[60] T. S. Rao and M. M. Gabr, "The estimation of the bispectral density function and the detection of periodicities in a signal," *Multivariate Statistics and Probability*, pp. 484–503, 1988. also printed in Journal of Multivariate Analysis Volume 27 No 2.

[61] P. J. Huber, B. Kleiner, T. Gasser, and G. Dummeruth, "Statistical methods for investigating phase relations in stationary stochastic processes," *IEEE Transactions on Audio and Electroacoustics*, vol. AU-19, no. 1, pp. 78–86, 1971.

[62] M. B. Priestley, *Nonlinear and Nonstationary time series analysis.* London: Academic Press, 1988.

[63] C. L. Nikias and A. P. Petropulu, *Higher-Order Spectra analysis.* PTR Prentice Hall, New Jersey, 1st ed., 1993.

[64] D. R. Brillinger, *Time Series : Data Analysis and Theory.* New York: Holt, Rinehart and Winston Inc, 1975.

[65] J. W. Dalle-Molle, *Higher-order spectral analysis and the trispectrum.* PhD thesis, The University of Texas at Austin, August 1992.

[66] W. B. Collis, *Higher order spectra and their application to nonlinear mechanical systems*. PhD thesis, University of Southampton, February 1996.

[67] M. J. Hinich and M. A. Wolinsky, "A test for aliasing using bispectral analysis," *Journal of the American Statistical Association*, vol. 83, no. 402, pp. 499–502, 1988.

[68] J. G. Proakis, C. M. Rader, F. Ling, and C. L. Nikias, *Advanced Digital Signal Processing*. New York: Macmillan, 1992.

[69] S. L. Marple, *Digital spectral analysis*. New Jersey: Prentice-Hall, 1987.

[70] P. R. White, July 1996. Private communication.

[71] J. W. A. Fackrell, S. McLaughlin, and P. R. White, "Bicoherence estimation using the direct method : Part 1 - theoretical considerations," *Applied Signal Processing*, 1996. (accepted for publication).

[72] S. Elgar and G. Sebert, "Statistics of bicoherence and biphase," *Journal of Geophysical Research*, vol. C94, pp. 10993–10998, 1989.

[73] D. Brillinger and M. Rosenblatt, "Asymptotic theory of estimates of k-th order spectra," in *Spectral Analysis of Time Signals* (B. Harris, ed.), pp. 153–188, Wiley, 1967.

[74] M. J. Hinich, "Testing for Gaussianity and linearity of a stationary time series," *Journal of Time Series Analysis*, vol. 3, no. 3, pp. 169–176, 1982.

[75] V. Chandran and S. L. Elgar, "Mean and variance of estimates of the bispectrum of a harmonic random process - an analysis including leakage effects," *IEEE Transactions on Signal Processing*, vol. 39, pp. 2640–2651, December 1991.

[76] S. Elgar and R. T. Guza, "Statistics of bicoherence," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, pp. 1667–1668, October 1988.

[77] T. Söderström, *Discrete-time Stochastic Systems : Estimation and Control*. Series in Systems and Control Engineering, New York: Prentice Hall, 1994.

[78] C. L. Nikias, "ARMA bispectrum approach to nonminimum phase system identification," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, no. 4, pp. 513–525, 1988.

[79] P. R. White, 1994. Private communication.

[80] C. L. Nikias and J. M. Mendel, "Signal processing with Higher Order Spectra," *IEEE Signal Processing Magazine*, pp. 10–36, July 1993.

[81] V. Kravtchenko-Berejnoi, *Polyspectral analysis and turbulent processes in the space plasmas*. PhD thesis, Laboratoire de Physique et Chimie de l'environment, University of Orléans, France, 1994.

[82] A. R. Lyons, T. J. Newton, N. J. Goddard, and A. T. Parsons, "Can passive sonar signals be classified on the basis of their higher order statistics ?," in *IEE Colloquium on 'Higher Order Statistics in Signal Processing : Are they of any use ?' Digest 1995/111*, (London, UK), pp. 6/1–6/6, May 1995.

[83] M. R. Raghuveer, "Higher-order statistics : Laying a myth to rest," in *Proceedings of the 29th Asilomar Conference on Signals, Systems and Computers*, (Pacific Grove, California, USA), pp. 5–8, 1995.

[84] G. M. Jenkins and D. G. Watts, *Spectral Analysis and its Applications*. San Fransisco: Holden-Day, 1968.

[85] M. J. Hinich and M. A. Wolinsky, "A test for aliasing using bispectral components," *Journal of the American Statistical Association Theory and Methods*, vol. 83, no. 402, pp. 499–502, 1993.

[86] W. B. Collis and P. R. White, "Bispectrum and trispectrum of mechanical systems." ISVR Internal Report, December 1994.

[87] G. Sebert and S. Elgar, "Statistics of bicoherence and biphase," in *Signal Processing Workshop on Higher-Order Spectra*, (Vail, Colorado, USA), pp. 223–228, IEEE, June 1989.

[88] J. W. A. Fackrell and S. McLaughlin, "Determining the false-alarm performance of HOS-based quadratic phase coupling detectors," in *Proceedings of EUSIPCO-96, Eighth European Signal Processing Conference*, (Trieste, Italy), September 1996. (to be presented).

[89] J. W. A. Fackrell, P. R. White, J. K. Hammond, R. J. Pinnington, and A. T. Parsons, "The interpretation of the bispectra of vibration signals: part 1 - theory," *Mechanical Systems and Signal Processing*, vol. 9, no. 3, pp. 257–266, 1995.

[90] J. W. A. Fackrell, P. R. White, J. K. Hammond, R. J. Pinnington, and A. T. Parsons, "The interpretation of the bispectra of vibration signals: part 2 - ex-

perimental results and applications," *Mechanical Systems and Signal Processing*, vol. 9, no. 3, pp. 267–274, 1995.

[91] E. Kreyszig, *Advanced Engineering Mathematics*. New York: John Wiley and Sons, 6th ed., 1988.

[92] H. Tong, *Non-linear time series : a dynamical system approach*. Oxford: Clarendon Press, 1990.

[93] M. J. Hinich, "Detecting a transient signal by bispectral analysis," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, no. 7, pp. 1277–1283, 1990.

[94] G. E. Ioup, J. W. Ioup, K. H. Barnes, R. L. Field, J. H. Leclere, and G. H. Rayborn, "Evaluation of bicorrelations for transient detection," in *Signal Processing Workshop on Higher-Order Spectra*, (Vail, Colorado, USA), pp. 46–51, 1989.

[95] L. A. Pflug, G. E. Ioup, J. W. Ioup, K. H. Barnes, R. L. Field, and G. H. Rayborn, "Detection of oscillatory and impulsive transients using higher order correlations and spectra," *Journal of the Acoustical Society of America*, vol. 91, pp. 2763–2776, May 1992.

[96] L. A. Pflug, G. E. Ioup, J. W. Ioup, and R. L. Field, "Properties of higher-order correlations and spectra for bandlimited deterministic transients," *Journal of the Acoustical Society of America*, vol. 91, pp. 975–988, February 1992.

[97] A. K. Nandi, "Robust estimation of third-order cumulants in applications of higher-order statistics," *IEE PROCEEDINGS F. Radar and Signal Processing*, vol. 6, pp. 380–389, December 1993.

[98] A. G. Stogioglou, 1994. Private communication.

[99] C. K. Papadopoulos and C. L. Nikias, "Bispectrum estimation of transient signals," in *International Conference on Acoustics, Speech and Signal Processing*, pp. 2404–2407, 1988.

[100] J. L. Rosenberger and M. Gasko, "Comparing location estimators : Trimmed means, medians and trimean," in *Understanding Robust and Exploratory data analysis* (D. C. Hoaglin, F. Mosteller, and J. W. Tukey, eds.), pp. 297–338, New York: John Wiley, 1983.

[101] J. W. A. Fackrell, A. G. Stogioglou, and S. McLaughlin, "Robust frequency-domain bicoherence estimation," in *IEEE Signal Processing Workshop on Statistical Signal and Array Processing*, (Corfu, Greece), pp. 206–209, 1996.

[102] V. Barnett and T. Lewis, *Outliers in Statistical Data*. Chichester: Wiley, 1994.

[103] F. J. Owens, *Signal processing of speech*. Macmillan, London, 1993.

[104] S. Hovell and B. Mulgrew, "Nonlinear analysis of drum sounds using higher order spectra," in *Proceedings of EUSIPCO-92, Sixth European Signal Processing Conference*, (Brussel, Belgium), pp. 1701–1704,, 1992.

[105] J. W. A. Fackrell, "Higher order spectral content of mechanical systems," Tech. Rep. ISVR Contract Report No 93/94, Institute of Sound and Vibration Research, University of Southampton, September 1993.

[106] A. Jongman and J. A. Sereno, "Acoustic properties of non-sibilant fricatives," in *International Conference on Phonetic Science*, vol. 4, (Stockholm, Sweden), pp. 432–435, 1995.

[107] B. F. J. Manly, *Multivariate Statistical Methods - A Primer*. London: Chapman and Hall, 2nd ed., 1994.

[108] M. J. Hinich and H. Messer, "On the principal domain of the disrcete bispectrum of a stationary signal," *IEEE Transactions on Signal Processing*, vol. 43, no. 9, pp. 2130–2134, 1995.

[109] O. J. Dunn and V. A. Clark, *Applied Statistics : Analysis of Variance and regression*. New York: John Wiley and Sons Inc, 1987.

[110] M. J. Norusis, *SPSS Base System User's Guide*. Chicago: SPSS Inc, 1990.

[111] M. L. Williams, *The use of the bispectrum and other higher order statistics in the analysis of one dimensional signals*. PhD thesis, Imperial College of Science, Technology and Medicine, University of London, UK, July 1992.

[112] G. Frazer, A. Reilly, and B. Boashash, "The bispectral aliasing test," in *IEEE Signal Processing Workshop on Higher-Order Statistics*, (Lake Tahoe, California, USA), pp. 332–335, IEEE, June 1993.

[113] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*. New York: Dover Publications Inc, 1972.

[114] R. D. Strum and D. E. Kirk, *First Principles of Discrete Systems and Digital Signal Processing*. Reading, Massachusetts: Addison-Wesley, 1989.

[115] J. W. A. Fackrell, S. McLaughlin, and P. R. White, "Bicoherence estimation using the direct method : Part 2 - practical considerations," *Applied Signal Processing*, 1996. (accepted for publication).

[116] M. Schmidt, 1994. Private communication.

[117] Audio-Technica US Inc, Ohio, USA, *Technical Specifications of ATM73a Head-worn cardiod condensor microphone*, 1989. Form No 0305-0727-01.

[118] Brüel and Kjær, "Condenser microphones and microphone preamplifiers for acoustic measurements," tech. rep., DK-2850 N/AE RUM, Denmark, 1982.

[119] C. H. Shadle, October 1995. Private communication.

[120] C. H. Shadle, P. Badin, and A. Moulinier, "Toward the spectral characteristics of fricative consonants," in *Proc Int Conf Phonetic Science*, pp. 42–45, August 1991.

Appendix A

Higher Order Statistics

A.1 The Higher Order Statistics of Gaussian signals

In this section it is demonstrated that the cumulants of a signal with a Gaussian pdf are identically zero. This is discussed in many texts on the subject of HOS (e.g. [62]) but rarely explicitly explained. It is included here since the result provides some insight into why HOS methods have appeal from a noise-rejection viewpoint.

The probability density function of a continuous Gaussian process of mean μ and variance σ^2 is [62]

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-(x-\mu)^2/(2\sigma^2)]$$

To see what the cumulants of this process are, first the Moment Generating Function (MGF) is calculated, from which the Cumulant Generating Function (CGF) and hence the cumulants can be determined.

The MGF $M(t)$ is simply the two-sided Laplace Transform of the pdf [62, p56]

$$M(t) = \exp[\mu t + \frac{1}{2}\sigma^2 t^2]$$

from which it follows that the CGF $K(t)$ $(= \ln M(t))$ is simply

$$K(t) = \mu t + \frac{1}{2}\sigma^2 t^2 \tag{A.1}$$

Now from the CGF the joint cumulants are determined by forming a power expansion of the form [62, p58]

$$K(t) = c_1 t + c_2 \frac{t^2}{2!} + c_3 \frac{t^3}{3!} + ...c_r \frac{t^2}{r!} + .. \tag{A.2}$$

where c_1 is the 1st cumulant of the process, c_2 the second and so on. Comparing Equations A.1 and A.2 it is evident that for a Gaussian process

$$\begin{aligned} c_1 &= \mu \\ c_2 &= \sigma^2 \\ c_r &= 0 \;\; r \geq 3. \end{aligned}$$

It thus follows that the zeroth-lag cumulants of order 3 and higher, that is $c_3(0,0)$, $c_4(0,0,0)$, and so on, are identically zero for a Gaussian process. To generalise this result to non-zero lags, a short heuristic argument will be given. Consider first the autocorrelation function $R(\tau)$ of a zero-mean process. The autocorrelation of this process is equal to the second-order cumulant $c_2(\tau)$. Now it is easy to show that $c_2(\tau)$ has its maximum possible value at $\tau = 0$, so that

$$c_2(\tau) \leq c_2(0) \;\; \forall \tau.$$

It thus follows directly, that if $c_2(0) = 0$ then $c_2(\tau) = 0 \, \forall \tau$. A similar argument can be applied to the third-order cumulant, which results in the conclusion that if $c_3(0,0) = 0$, then $c_3(\tau_1, \tau_2) = 0 \, \forall \tau_1, \tau_2$. Since $c_3(0,0)$ is just the signal skewness, this means that signals with zero skew have zero third-order cumulants.

A.2 Test for "Gaussianity"

This section describes some standard properties of the skewness function $s^2(k,l)$ which have been proposed [74] to test whether or not a signal is Gaussian. These results are important because in Chapter 4 they are used as the first part of a novel QPC detector. Some simple manipulations are also given which allow the properties of the skewness function to be translated into critical levels for significant squared bicoherence magnitude (as used in Chapter 4), and for critical levels for squared bicoherence averaged over the IT (as used in Chapter 7).

A.2.1 Determining critical levels for each bifrequency

Several researchers [60, 74] have put the result of Property 3 (in Section 3.6.1) into a hypothesis testing framework. Under the null hypothesis H_0 that the signal of interest is Gaussian, it has been shown (see for example, [74]) that the scaled skewness function $2Ks^2(k,l)$ is asymptotically centrally-χ^2 distributed with 2 degrees of freedom (dof). This result can be used to derive approximate critical levels for significant squared bicoherence.

Let $c_{\alpha(1)}$ be the critical level with significance $\alpha(1)$ for a central-χ^2 distribution with 2 dof (from standard tables). Under H_0 Hinich's results [74] imply that

$$P\left(2Ks^2(k,l) > c_{\alpha(1)}\right) = \alpha(1)$$

from which it follows directly that

$$P\left(s^2(k,l) > c_{\alpha(1)}/2K\right) = \alpha(1)$$

and, because in this case the bicoherence function has been found to have the same statistical properties as the skewness function [72], $b^2(k,l)$ can be substituted for $s^2(k,l)$ to give

$$P\left(b^2(k,l) > c_{\alpha(1)}/2K\right) = \alpha(1),$$

Under H_0 (that the signal is Gaussian) the contour level for significant values of the squared bicoherence is $c_{\alpha(1)}/2K$.

Thus if the squared bicoherence at a particular bifrequency (k,l) is greater than the threshold $c_{\alpha(1)}/2K$, then the null hypothesis, that the signal is Gaussian noise, is rejected, and the alternative hypothesis, that the signal is non-Gaussian, is accepted.

An approach similar to this was taken in [111]. The good property that this test has is that it produces a result for each bifrequency, and so it provides detailed information about where in the bifrequency plane interesting signal properties occur. However, care is required because if this test is simultaneously applied to many bifrequencies, for example by applying a threshold of $c_{\alpha(1)}/2K$ to the whole IT, then the probabilities of false detections can accumulate, resulting in an overestimate of the number of bifrequencies at which the magnitude is significant.

A.2.2 Determining critical levels for summations over bifrequencies

An estimator with better statistical properties, but no frequency resolution, is formed by summing $s^2(k,l)$, or equivalently $b^2(k,l)$, over one of the triangles in the PD. Here only the summation over the IT is considered, as this relates to the Gaussian hypothesis. For details of the properties of the sum over the OT, which may hold information about aliasing and stationarity (this matter is under some dispute [112]) see [67, 108].

Following the result stated in Section A.2.1 above, it follows that under H_0 (that the signal is Gaussian), the summation of the skewness function over the L bifrequencies in the IT is asymptotically centrally-χ^2 distributed, but now with $2L$ dof. A hypothesis test can thus be carried out, comparing the skewness function estimate $2K \sum_{\mathrm{IT}} \hat{s}^2(k,l)$ with a critical value $c_\alpha^{\chi^2}$ from standard central-χ^2 tables. The reason for summing the skewness function over the IT is that this quantity is proportional to $\gamma_3 = c_3(0,0)$, the *skewness* of the signal[1]. Again using the result that b^2 and s^2 are approximately equal, this gives

> Under H_0 (that the signal is Gaussian) the critical level for significant values of the squared bicoherence b^2 summed over the L bifrequencies in the IT is $c_{\alpha(1)}^{\chi^2}/2K$ where $c_{\alpha(1)}^{\chi^2}$ is determined from central-χ^2 tables with $2L$ dof.

If the number of bifrequencies in the IT is large (e.g. if it is more than 100) then a more convenient test can be formulated[65] by using a normal approximation to the central-χ^2 distribution [113, p941]

$$Z = \sqrt{2\chi^2} - \sqrt{2(\mathrm{dof}) - 1}, \tag{A.3}$$

in which Z is a standardised normal variable N(0,1).

A.2.3 The averaged squared bicoherence

In addition to the two types of test for Gaussian signals described above, in this section a further measure is introduced - $\overline{b^2_{\mathrm{IT}}}$ - the squared bicoherence *averaged* over the IT, as this is used in Chapter 7. In this thesis, the quantity $\overline{b^2_{\mathrm{IT}}}$ is mostly interpreted as

[1] This is why $s^2(k,l)$ is called the "*skewness* function".

an indicator of the average level of QPC, but it can also be interpreted in the "Gaussianity" hypothesis-testing framework described above. In this section, the relation is established[2] between the critical value of $\sum_{\mathrm{IT}} b^2$ for detecting non-Gaussian signals (from Section A.2.2 above and [74]) and the squared bicoherence averaged over the IT, $\overline{b^2_{\mathrm{IT}}}$. This results in an expression for a critical level of $\overline{b^2_{\mathrm{IT}}}$ above which the null hypothesis that the signal is Gaussian is rejected.

The critical level for $\overline{b^2_{\mathrm{IT}}}$ is determined in the following way :

From Section A.2 it is known that under the null hypothesis H_0 (that the signal is Gaussian) $2K \sum_{\mathrm{IT}} b^2$ is centrally-χ^2 distributed with $2L$ dof.

If the squared bicoherence is *averaged* over the IT to give $\overline{b^2_{\mathrm{IT}}}$,

$$\overline{b^2_{\mathrm{IT}}} \triangleq \frac{1}{L} \sum_{\mathrm{IT}} b^2,$$

then it follows trivially that $L \times 2K\overline{b^2_{\mathrm{IT}}}$ is also centrally-χ^2 distributed with $2L$ dof (since $L\overline{b^2_{\mathrm{IT}}} \equiv \sum_{\mathrm{IT}} b^2$). All the bispectral analysis discussed in this thesis involves IT sizes of at least 225 bins (for $M = 64$), and so the normal approximation described in Equation A.3 can be used.

If the critical value for Z with significance level $\alpha(1)$ is denoted $c^z_{\alpha(1)}$, and that for the χ^2 variable (with $2L$ dof) as $c^{\chi^2(2L)}_{\alpha(1)}$ then Equation A.3 can be written

$$\begin{aligned} \sqrt{2c^{\chi^2(2L)}_{\alpha(1)}} &= c^z_{\alpha(1)} + \sqrt{4L-1} \\ \Rightarrow c^{\chi^2(2L)}_{\alpha(1)} &= \frac{1}{2}\left[c^z_{\alpha(1)} + \sqrt{(4L-1)}\right]^2 \end{aligned}$$

Thus the critical level for the estimate $2KL\overline{b^2_{\mathrm{IT}}}$ (which is $\chi^2(2L)$ under H_0) is

$$\mathrm{Critical}[2KL\overline{b^2_{\mathrm{IT}}}] = \frac{1}{2}\left[c^z_{\alpha(1)} + \sqrt{(4L-1)}\right]^2,$$

from which it easily follows that the critical level for $\overline{b^2_{\mathrm{IT}}}$ is

$$\mathrm{Critical}\left[\overline{b^2_{\mathrm{IT}}}\right] = \frac{1}{4KL}\left[c^z_{\alpha(1)} + \sqrt{(4L-1)}\right]^2,$$

and thus:

[2]The indices (k, l) will be dropped in this section for clarity.

Under H_0 (that the signal is Gaussian) the critical level for significant values of $\overline{b^2_{\rm IT}}$, the squared bicoherence b^2 *averaged* over the L bifrequencies in the IT, is approximately $\frac{1}{4KL}\left[c^z_{\alpha(1)} + \sqrt{(4L-1)}\right]^2$, where $c^z_{\alpha(1)}$ is a one-sided critical value from the standard normal distribution.

The critical values for the standard normal distribution[3], and the critical values of $\overline{b^2_{\rm IT}}$ corresponding to them for a $K = 64$ bispectral analysis, are shown in Table A.1. The critical value chosen for use in most of the analysis in this thesis is at the $\alpha = 0.001$ level, which corresponds to Critical$[\overline{b^2_{\rm IT}}] = 0.0190$.

Sig. level α	c_z : $P(Z > c^z_\alpha) = \alpha$	Critical threshold for $\overline{b^2_{\rm IT}}$
.05	1.645	.0174
.01	2.326	.0181
.005	2.576	.0184
.001	3.090	.0190

Table A.1: Some critical values of $\overline{b^2_{\rm IT}}$ for $K = 64$.

A.3 The effect of Gaussian noise on the skewness function

In this section it is shown that the skewness function $s^2(k,l)$ (Equation 3.10, Section 3.4.4.1) is not "blind" to Gaussian noise.

Consider a signal formed by adding together a non-Gaussian signal and a Gaussian signal, as shown in Figure A.1. Assume that the two signals are independent of each other, so that when they are combined no cross terms arise.

In theory the bispectrum is "blind" to Gaussian noise (from Property 1, Section 3.6), and so the bispectrum of $x(n)$ will be the same as the bispectrum of the non-Gaussian signal $u(n)$.

$$\begin{aligned} E[B_x(k,l)] &= E[B_u(k,l)] + E[B_v(k,l)] \\ &= E[B_u(k,l)]. \end{aligned} \tag{A.4}$$

[3] Critical values corresponding to one-sided probabilities $P(Z > c^z_{\alpha(1)}) = \alpha$ are used.

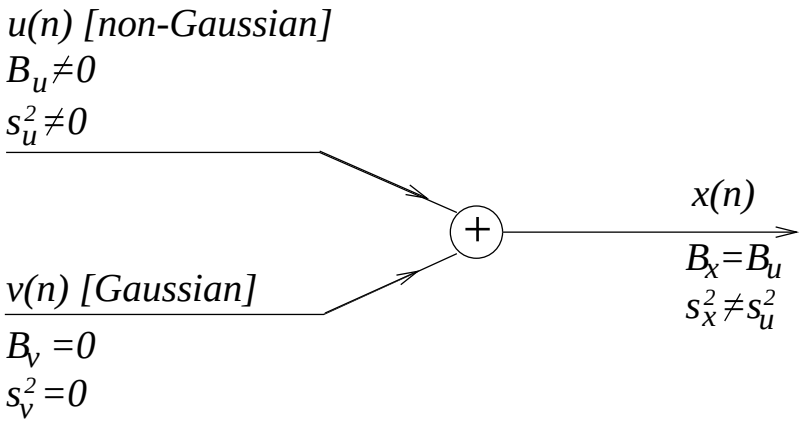

Figure A.1: Schematic diagram showing signal formed by adding Gaussian and non-Gaussian components.

Thus it follows that the numerator of the skewness function (Equation 3.10) will be unchanged by additive Gaussian noise.

However, the denominator of Equation 3.10 **will** change, since the power spectrum of $x(n)$ depends on the spectra of *both* the Gaussian and non-Gaussian components;

$$E[P_x(k)] = E[P_u(k)] + E[P_v(k)]$$

from which it follows easily that,

$$\begin{aligned} E[P_x(k)]E[P_x(l)]E[P_x(k+l)] &= (E[P_u(k)] + E[P_v(k)]) \times \\ &\quad (E[P_u(l)] + E[P_v(l)]) \times \\ &\quad (E[P_u(k+l)] + E[P_v(k+l)]) \\ &= E[P_v(k)]E[P_v(l)]E[P_v(k+l)] + \\ &\quad E[P_u(k)]E[P_u(l)]E[P_u(k+l)] + \\ &\quad [6 \text{ cross terms}]. \end{aligned}$$

Then it follows that the denominator of the skewness function of $x(n)$ is *greater than or equal to* the denominator of the skewness function of $u(n)$, even though their bispectra are the same;

$$E[P_x(k)]E[P_x(l)]E[P_x(k+l)] \geq E[P_u(k)]E[P_u(l)]E[P_u(k+l)], \qquad \text{(A.5)}$$

where the equality only holds in the case where there is a zero in the noise spectrum $P_v(k)$ (a condition which in practice is very unlikely to happen).

Finally substituting Equations A.4 and A.5 into Equation 3.10 it becomes clear that $s_x^2(k,l) \leq s_u^2(k,l)$, that is, that by adding Gaussian noise to the non-Gaussian signal, the skewness function is reduced. Hence the skewness function is not "blind" to Gaussian noise.

A.4 Filtering Effects

Normalised and unnormalised bispectra have interesting filter-invariance properties which need to be understood before applying bispectral analysis. This section considers some issues related to filtering, and its effect on bispectral measures.

Consider the filtering operation shown in Figure A.2. The filter is assumed to be *linear*, causal and time-invariant. This operation can be represented as a convolution in the time domain (such as Equation 3.3)

$$y(n) = \sum_{m=0}^{\infty} h(m)x(n-m),$$

or as a multiplication in the frequency domain, relating the FT's of the input and output signals via the *transfer function* of the filter

$$Y(k) = H(k)X(k) \tag{A.6}$$

A.4.1 Linear Filters

From Equation 3.10 the skewness function can be rewritten in terms of Fourier Transforms

$$\begin{aligned} s_y^2(k,l) &\triangleq \frac{|E[B(k,l)]|^2}{E[P(k)]E[P(l)]E[P(k+l)]} \\ &= \frac{|E[Y(k)Y(l)Y^*(k+l)]|^2}{E[Y(k)Y^*(k)]E[Y(l)Y^*(l)]E[Y(k+l)Y^*(k+l)]}. \end{aligned}$$

Now if each $Y()$ is replaced by $H(k)X(k)$, then the $H()$ terms can be taken *outside* the expectation operator (since the filter is time invariant), from which it follows that

$$s_y^2(k,l) = \frac{|H(k)H(l)H^*(k+l)|^2}{H(k)H^*(k)H(l)H^*(l)H(k+l)H^*(k+l)} \times \frac{|E[X(k)X(l)X^*(k+l)]|^2}{E[X(k)X^*(k)]E[X(l)X^*(l)]E[X(k+l)X^*(k+l)]}. \quad \text{(A.7)}$$

Now the $H()$ terms cancel, giving the result

$$\boxed{s_x^2(k,l) = s_y^2(k,l).}$$

From a practical point of view, this only holds if the condition $H(k) > 0 \; \forall \; k$ is met (this prevents terms of the type 0/0 occurring in Equation A.7). This is equivalent to requiring that the linear filter does not have any zeros on the unit circle.

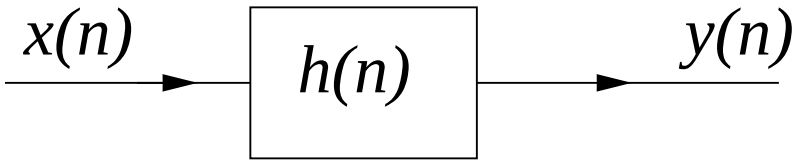

Figure A.2: Schematic diagram showing simple linear filter.

A.4.2 Linear Phase Filters

Although the normalised bispectral quantities such as skewness and bicoherence are magnitude-only quantities, and so have no phase, under certain conditions (explored in Chapter 4) the phase of the bispectrum is of interest. In this section the effect of filters on the bispectrum phase (or *biphase*) is investigated.

Consider the simple linear filter shown schematically in Figure A.2. Equation A.6 can be written in a slightly more explicit way by expanding each transfer function into a magnitude and phase component;

$$\begin{aligned} Y(k) = |Y(k)|e^{j\phi_y(k)} &= |H(k)|e^{j\phi_h(k)}|X(k)|e^{j\phi_x(k)} \\ &= |H(k)||X(k)|e^{j(\phi_h(k)+\phi_x(k))}, \end{aligned}$$

from which the bispectrum is given by

$$\begin{aligned} B_y(k,l) &= E\left[Y(k)Y(l)Y^*(k+l)\right] \\ &= E\left[H(k)X(k)\times H(l)X(l)\times H^*(k+l)X^*(k+l)\right] \\ &= E\left[|H(k)||X(k)||H(l)||X(l)||H^*(k+l)||X^*(k+l)| \right. \\ &\quad \left. e^{j(\phi_x(k)+\phi_x(l)-\phi_x(k+l)+\phi_h(k)+\phi_h(l)-\phi_h(k+l))}\right]. \end{aligned}$$

It is evident that for the filter to preserve the phase characteristics of the bispectrum of the input signal, then $\angle B_y(k,l) = \angle B_x(k,l)$, and this is possible if

$$\angle B_h(k,l) = \phi_h(k) + \phi_h(l) - \phi_h(k+l) = 0, \pm 2\pi, \pm 4\pi, ... \tag{A.8}$$

A class of filters which have this property are the *linear phase filters*, characterised by the phase characteristic [114]:

$$\phi_h(k) = -\alpha k$$

Substitution of this phase characteristic into Equation A.8 gives

$$\angle B_h(k,l) = -\alpha k - \alpha l + \alpha(k+l) = 0.$$

Thus linear phase filters preserve biphase information.

A.5 Theoretical Bicoherence of Harmonic Signals

In this section the theoretical squared bicoherence is derived for a harmonics in noise problem. The approach taken here follows closely that used elsewhere in a consideration of (unnormalised) bispectra [75], extending the analysis to normalised bispectra, and also considering the effect of noise. The effects of windowing and noise are treated separately in Sections A.5.1 and A.5.2 respectively.

A.5.1 Windowing Effects

In each case the signal is assumed to conform to the M2 model from Equation 3.4 in Section 3.4.1 but the phases and amplitudes are now defined to be random variables,

which means that they are described by their statistical properties[4]. The model is given by

$$x(n) = \sum_{p=1}^{P} A_p \cos(2\pi f_p n + \phi_p).$$

and extensive use will be made of the Dirichlet kernel

$$D(k, f_p, M) = \frac{\sin\{\pi(k - f_p M)\}}{M \sin\{\pi(k - f_p M)/M\}}.$$

A.5.1.1 The Power Spectrum

Chandran [75, eqn 7] gives the DFT coefficients $X(k)$ as

$$\begin{aligned} X(k) &= \sum_{p=1}^{P} \frac{A_p}{2} \Big[e^{-j\pi(M-1)(k/M - f_p)} D(k, f_p, M) e^{j\phi_p} \\ &\quad + e^{-j\pi(M-1)(k/M + f_p)} D(k, -f_p, M) e^{-j\phi_p} \Big] \end{aligned}$$

The periodogram-averaged power spectrum estimate $\hat{P}(k)$ is then given by

$$\begin{aligned} E[\hat{P}(k)] &= E\left[\frac{1}{K}\sum_{i=1}^{K} X_i(k) X_i^*(k)\right] \\ &= \frac{1}{K}\sum_{i=1}^{K} E\left[X_i(k) X_i^*(k)\right] \quad \text{(linearity)} \\ &= E\left[X_i(k) X_i^*(k)\right] \\ &= E\left[X(k) X^*(k)\right] \quad \text{(stationarity)} \\ &= \sum_{p=1}^{P}\sum_{q=1}^{P} E\left[\frac{A_p A_q}{4} \right. (\\ &\quad e^{-j\pi(M-1)(-f_p + f_q)} e^{j\phi_p - \phi_q} D(k, f_p, M) D(k, f_q, M) \\ &\quad + e^{-j\pi(M-1)(-f_p - f_q)} e^{j\phi_p + \phi_q} D(k, f_p, M) D(k, -f_q, M) \\ &\quad + e^{-j\pi(M-1)(f_p + f_q)} e^{j - \phi_p - \phi_q} D(k, -f_p, M) D(k, f_q, M) \\ &\quad \left. + e^{-j\pi(M-1)(-f_p + f_q)} e^{j\phi_p - \phi_q} D(k, -f_p, M) D(k, -f_q, M) \Big) \right]. \end{aligned}$$

[4] This approach follows that of [75], but if the sinusoid amplitudes are instead defined as deterministic constants, then it is simple to show that in fact the same results arise.

Now rearranging this last expression gives

$$
\begin{aligned}
E[\hat{P}(k)] \;=\; & \sum_{p=1}^{P}\sum_{q=1}^{P} E\left[\frac{A_p A_q}{4}(\right. \\
& e^{-j\pi(M-1)(-f_p-f_q)} e^{j\phi_p+\phi_q} D(k,f_p,M) D(k,-f_q,M) \\
& + e^{-j\pi(M-1)(f_p+f_q)} e^{j-\phi_p-\phi_q} D(k,-f_p,M) D(k,f_q,M) \\
& + e^{-j\pi(M-1)(-f_p+f_q)} e^{j\phi_p-\phi_q} D(k,f_p,M) D(k,f_q,M) \\
& \left. + e^{-j\pi(M-1)(-f_p+f_q)} e^{j\phi_p-\phi_q} D(k,-f_p,M) D(k,-f_q,M)\right] .
\end{aligned}
$$

To simplify this expression the following assumptions are made

A1 The statistics of the sinusoid phases are independent of the statistics of the sinusoid amplitudes. This means the expectation of terms involving these quantities can be considered separately.

A2 The sinusoids are statistically independent. This means that for the cases where $p \neq q$ the expectations of the form $E[e^{j(\pm\phi_p \pm \phi_q)}]$ can be rewritten $E[e^{j(\pm\phi_p)}]E[e^{j(\pm\phi_q)}]$, but only if $p \neq q$.

A3 The sinusoid phases are Uniformly distributed $U[0, 2\pi)$. This then means that $E[e^{j(\pm\phi_p)}] = 0 \;\; p = 1, .., P$.

Now the above equation can be rewritten, splitting the RHS into two parts - those for which $p \neq q$ and those for which $p = q$. In the latter cases the double summation $\sum_{p=1}^{P}\sum_{q=1,p\neq q}^{P}$ can be replaced by a single summation $\sum_{p=1}^{P}$. The equation for the

power spectral estimate is then

$$
\begin{aligned}
E[\hat{P}(k)] \;=\; & \sum_{p=1}^{P} \sum_{q=1,q\neq p}^{P} E\left[\frac{A_p A_q}{4}\right] \times \\
& \Big[e^{-j\pi(M-1)(-f_p-f_q)} E[e^{j\phi_p}]E[e^{j\phi_q}] D(k,f_p,M)\,D(k,-f_q,M) \\
& + e^{-j\pi(M-1)(f_p+f_q)} E[e^{-j\phi_p}]E[e^{-j\phi_q}] D(k,-f_p,M)\,D(k,f_q,M) \\
& + e^{-j\pi(M-1)(f_p-f_q)} E[e^{-j\phi_p}]E[e^{j\phi_q}] D(k,f_p,M)\,D(k,f_q,M) \\
& + e^{-j\pi(M-1)(-f_p+f_q)} E[e^{j\phi_p}]E[e^{-j\phi_q}] D(k,-f_p,M)\,D(k,-f_q,M)\Big] \\
& + \sum_{p=1}^{P} E\left[\frac{A_p^2}{4}\right] \times \\
& \Big[e^{-j\pi(M-1)(-f_p-f_p)} E[e^{j(\phi_p+\phi_p)}] D(k,f_p,M)\,D(k,-f_p,M) \\
& + e^{-j\pi(M-1)(f_p+f_p)} E[e^{j(-\phi_p-\phi_p)}] D(k,-f_p,M)\,D(k,f_p,M) \\
& + e^{-j\pi(M-1)(-f_p+f_p)} E[e^{j(\phi_p-\phi_p)}] D(k,f_p,M)\,D(k,f_p,M) \\
& + e^{-j\pi(M-1)(f_p-f_p)} E[e^{j(-\phi_p+\phi_p)}] D(k,-f_p,M)\,D(k,-f_p,M)\Big]. \quad \text{(A.9)}
\end{aligned}
$$

Now the first four terms on the RHS arise from the cases where $p \neq q$, and each contains an expression of the form $E[e^{\pm j\phi_p}]$ which is zero because of Assumption A2. Of the four terms arising from the case $p = q$ only the last two shown have nonzero expectations. So Equation A.9 reduces to

$$
\begin{aligned}
E[\hat{P}(k)] \;=\; & \sum_{p=1}^{P} E\left[\frac{A_p^2}{4}\right] \\
& \Big[e^{-j\pi(M-1)(-f_p+f_p)} e^{j(\phi_p-\phi_p)} D(k,f_p,M)\,D(k,f_p,M) \\
& + e^{-j\pi(M-1)(f_p-f_p)} e^{j(-\phi_p+\phi_p)} D(k,-f_p,M)\,D(k,-f_p,M)\Big]. \quad \text{(A.10)}
\end{aligned}
$$

It is easy to see that the exponential terms are all $e^0 = 1$ and so the power spectral estimate is

$$
E[\hat{P}(k)] \;=\; \sum_{p=1}^{P} \frac{E[A_p^2]}{4} \Big[D^2(k,f_p,M) + D^2(k,-f_p,M)\Big]. \quad \text{(A.11)}
$$

A.5.1.2 The Bispectrum

Previous researchers [75] have used the above method to derive an expression for the bispectrum of the same signal:

$$
\begin{aligned}
E[\hat{B}(k,l)] \;=\; & \sum_{p,q,r,=1}^{P} \frac{E[A_p A_q A_r]}{8} \Bigg[\sum_{a,b,c=0}^{1} \\
& D(k,(-1)^a f_p, M) D(l,(-1)^b f_q, M) D(k+l,(-1)^{c+1} f_r, M) \times \\
& e^{-j\pi(M-1)((-1)^a f_p + (-1)^b f_q + (-1)^c f_r)} \times \\
& E[e^{j((-1)^a \phi_p + (-1)^b \phi_q + (-1)^c \phi_r)}] \Bigg] . \qquad \text{(A.12)}
\end{aligned}
$$

A.5.1.3 The skewness

Although the procedure adopted so far has been to use the bicoherence, rather than the skewness function, for analysis of M2 signals, the analysis which follows is made clearer by considering the skewness function. In fact it can be shown that the same final result holds for the bicoherence also.

By combining equations A.11 and A.12 an expression can be obtained for the skewness function (Equation 3.12) in terms of the model parameters. The numerator of this expression is the modulus squared of equation A.12

$$
\begin{aligned}
| E[\hat{B}(k,l)] |^2 \;=\; & \Bigg| \sum_{p,q,r,=1}^{P} \frac{E[A_p A_q A_r]}{8} \Bigg[\sum_{a,b,c=0}^{1} \\
& D(k,(-1)^a f_p, M) D(l,(-1)^b f_q, M) D(k+l,(-1)^{c+1} f_r, M) \times \\
& e^{-j\pi(M-1)((-1)^a f_p + (-1)^b f_q + (-1)^c f_r)} E[e^{j((-1)^a \phi_p + (-1)^b \phi_q + (-1)^c \phi_r)}] \Bigg] \Bigg|^2
\end{aligned}
$$

and the denominator of the skewness function is simply

$$
\begin{aligned}
\hat{P}(k)\hat{P}(l)\hat{P}(k+l) \;=\; & \sum_{p,q,r=1}^{P} \frac{E[A_p^2]}{4} \frac{E[A_q^2]}{4} \frac{E[A_r^2]}{4} \\
& \times \Big[D^2(k, f_p, M) + D^2(k, -f_q, M) \Big] \\
& \times \Big[D^2(l, f_q, M) + D^2(l, -f_p, M) \Big] \\
& \times \Big[D^2(k+l, f_r, M) + D^2(k+l, -f_r, M) \Big] \qquad \text{(A.13)}
\end{aligned}
$$

Now for the no-leakage case Equation A.13 reduces to

$$\begin{aligned} |\, E[\hat{B}(k,l)]\,|^2 &= |\, \frac{E[A_p A_q A_r]}{8}\,|^2 \text{ if } (k,l) = (Mf_1, Mf_2), (Mf_2, Mf_1) \\ &= 0 \text{ otherwise,} \end{aligned}$$

where the normalised frequencies f_1 and f_2 lead to bispectral content at the discrete frequencies Mf_1 and Mf_2 respectively. The denominator (Equation A.10) becomes

$$\hat{P}(k)\hat{P}(l)\hat{P}(k+l) = E[\frac{A_p^2 A_q^2 A_r^2}{64}]$$

and so the skewness function is

$$\begin{aligned} s^2(k,l) &= 1 \text{ if} (k,l) = (Mf_1, Mf_2), (Mf_2, Mf_1) \\ &= 0/0 \text{ otherwise.} \end{aligned}$$

The way in which the $0/0$ is dealt with can have a big effect on the skewness function output [115]. A simple way to deal with this problem is to add a small positive constant to the denominator of the skewness (or bicoherence) estimator. This is analogous to adding low level white Gaussian noise to the signal - it ensures that there is always some energy in each frequency bin, and so even if the bispectrum is zero, the denominator is always greater than zero, and so the $0/0$ problem does not arise.

If there is leakage then the bispectrum can be nonzero over the whole bispectral plane [75], and as a result of this the skewness function is also nonzero everywhere. For simulation signals, with no added noise, this can result in a skewness function which is unity everywhere. However, this situation changes radically if there is noise present[5], since the noise (which may arise from experimental measurement noise or from rounding errors) affect the numerator and denominator in different ways. Primarily this is because the variance of the numerator is higher than the variance of the denominator.

The situtation can be visualised as follows : the bispectrum of the clean signal has sidelobe effects leading to nonzero bispectra all across the bispectral plane. At any bifrequency, the ratio of the numerator to the denominator of $s^2(k,l)$ (and hence the value of the estimate of $s^2(k,l)$) is dependent on the SNR at the triplet of frequencies k, l and $k+l$. These effects, which can be viewed as an extension of the results in

[5] as opposed to a sinusoid-only signal.

Appendix A.3 are investigated in Appendix A.5.2 below, although the focus is on the squared bicoherence function b^2, since those results will prove useful in Chapter 4 also.

A.5.2 Noise effects

In this section the theoretical squared bicoherence, $b^2(k,l)$, is derived for an M2 signal $x(n)$ (see Equation 3.4 in Section 3.4.1) composed of sinusoids $y(n)$ (with $P = 3$) in additive noise $v(n)$, such that

$$\begin{aligned} y(n) &= \sum_{p=1}^{3} A_p \cos(2\pi f_p n + \phi_p) \\ x(n) &= y(n) + v(n). \end{aligned} \tag{A.14}$$

Windowing effects (treated in Section A.5.1) are ignored in this section so that the expressions do not become too complicated.

The bispectrum of $x(n)$ is determined by substituting expressions for its DFT $X(k) = Y(k)+V(k)$ into the expression for the squared bicoherence $b^2(k,l)$ (from Equation 3.13 in Section 3.4.4.2). The noise-free case is considered first.

A.5.2.1 Noise-free case

If there is no noise then $\sigma_v^2 = 0$ then it is trivial to show that $X(k) = Y(k)$, and that at discrete bifrequency corresponding to (f_1, f_2) the expected values of the components of the bicoherence (Eqn. 3.13) will be [71];

$$\begin{aligned} |E[X(f_1)X(f_2)X^*(f_3)]|^2 = |E[Y(f_1)Y(f_2)Y^*(f_3)]|^2 &= \prod_{i=1}^{3} \frac{A_i^2}{4}, \\ E[|X(f_1)X(f_2)|^2] = E[|Y(f_1)Y(f_2)|^2] &= \prod_{i=1}^{2} \frac{A_i^2}{4}, \\ E[|X(f_3)|^2] = E[|Y(f_3)|^2] &= \frac{A_3^2}{4}, \end{aligned}$$

from which it easily follows that if there is frequency coupling (so that $f_3 = f_1 + f_2$), then $b^2(f_1, f_2) = 1$.

A.5.2.2 Noisy case

If there is some additive noise $\sigma_v^2 \neq 0$ then the DFT of $x(n)$ becomes $X(k) = Y(k) + V(k)$. The numerator of the bicoherence is unchanged:

$$|E[X(f_1)X(f_2)X^*(f_3)]|^2 = |E[Y(f_1)Y(f_2)Y^*(f_3)]|^2 = \prod_{i=1}^{3} \frac{A_i^2}{4}, \tag{A.15}$$

since the cross terms between signal and noise which arise all have zero expectations, and the noise has zero expected bispectrum since it is assumed Gaussian. However, the two terms on the denominator do get affected by noise; the first term on the denominator becomes

$$\begin{aligned} E[|X(f_1)X(f_2)|^2] &= E[|(Y(f_1)+V(f_1))(Y(f_2)+V(f_2))|^2] \\ &= E[|Y(f_1)Y(f_2)+Y(f_1)V(f_2)+V(f_1)Y(f_2)+V(f_1)V(f_2)|^2] \\ &= E[Y(f_1)^2Y(f_2)^2+Y(f_1)^2V(f_2)^2+Y(f_2)^2V(f_1)^2] \\ &= E[Y(f_1)^2Y(f_2)^2]+\frac{\sigma_v^2}{M}\left(E[Y(f_1)^2]+E[Y(f_2)^2]\right) \\ &= \prod_{i=1}^{2}\frac{A_i^2}{4}+\frac{\sigma_v^2}{M}\sum_{i=1}^{2}\frac{A_i^2}{4}, \end{aligned} \tag{A.16}$$

where the independence of the noise $v(n)$ from the sinusoids has been used, and the fact that, for second-order white noise $E[V^2(k)] = \sigma_v^2/M$. The second term on the denominator becomes

$$E[|X(f_3)|^2] = E[|Y(f_3)+V(f_3)|^2] = E[Y(f_3)^2]+\frac{\sigma_v^2}{M} = \frac{A_3^2}{4}+\frac{\sigma_v^2}{M}. \tag{A.17}$$

The theoretical bicoherence $b^2(f_1, f_2)$ from Eqn. 3.13 is then given by combining Eqns. A.15-A.17, to give

$$\begin{aligned} b^2(f_1,f_2) &= \frac{\prod_{i=1}^{3}\frac{A_i^2}{4}}{\left[\prod_{i=1}^{2}\frac{A_i^2}{4}+\frac{\sigma_v^2}{M}\sum_{i=1}^{2}\frac{A_i^2}{4}\right]\left[\frac{A_3^2}{4}+\frac{\sigma_v^2}{M}\right]} \\ &= \frac{\prod_{i=1}^{3}A_i^2}{\prod_{i=1}^{3}A_i^2+\frac{4A_3^2\sigma_v^2}{M}\sum_{i=1}^{2}A_i^2+\frac{4\sigma^2}{M}\prod_{i=1}^{2}A_i^2+\frac{\sigma_v^4}{M^2}\sum_{i=1}^{2}A_i^2} \end{aligned} \tag{A.18}$$

It will now be assumed that the harmonic amplitudes are of equal amplitude $A_1 =$

$A_2 = A_3 = A$. This leads to the simplified expression

$$\begin{aligned} b^2(f_1, f_2) &= \frac{\frac{A^6}{64}}{\left[\frac{A^4}{16} + \frac{\sigma_v^2}{M}\frac{2A^2}{4}\right]\left[\frac{A^2}{4} + \frac{\sigma_v^2}{M}\right]} \\ &= \frac{1}{1 + \frac{12}{M}\frac{\sigma_v^2}{A^2} + \frac{32}{M^2}\frac{\sigma_v^4}{A^4}} \end{aligned} \tag{A.19}$$

Now the SNR is defined by

$$\text{SNR} \triangleq 10 \log_{10} \frac{\sigma_x^2}{\sigma_v^2}. \tag{A.20}$$

The variance of a real sine wave of amplitude A is $A^2/2$, and so Equation A.20 can be written

$$\begin{aligned} \text{SNR} &\triangleq 10 \log_{10} \frac{3A^2}{2\sigma_v^2} \\ \Rightarrow \quad \frac{\sigma^2}{A^2} &= \frac{3}{2} 10^{-\text{SNR}/10} \end{aligned}$$

which can be substituted into Equation A.19 to give

$$\boxed{b^2(f_1, f_2) = \frac{1}{1 + \frac{18}{M} 10^{-\text{SNR}/10} + \frac{72}{M^2} 10^{-2\text{SNR}/10}}.} \tag{A.21}$$

This equation indicates how the bicoherence peak changes as the SNR changes. If the SNR is very high then the second and third terms in the denominator will be very small and the bicoherence will be close to unity. As the SNR decreases, so the two denominator terms come into play, reducing the bicoherence peak.

Appendix B

Quadratic Phase Coupling

B.1 Probability of False Alarm for QPC detectors

In Section 4.5.2 it was shown that the P_{FA} for the biphase test could be evaluated using the relations

$$\begin{aligned} P_{FA} &= \frac{\theta_c}{\pi} \\ &\approx \frac{c_{\alpha(2)}}{2K\pi}\left(\frac{1}{b(k,l)^2}-1\right), \end{aligned}$$

where the approximation arises because the expression for the biphase variance used is approximate.

Additionally, in Appendix A.5.2 it was shown that, for 3 equal-amplitude sinusoids in AWGN, $b^2(k,l)$ is related to the SNR by

$$b^2(k,l) = \frac{1}{1+\frac{18}{M}10^{-\mathrm{SNR}/10}+\frac{72}{M^2}10^{-2\mathrm{SNR}/10}}.$$

Combining these expressions gives an estimate of the P_{FA} in terms of the SNR.

$$P_{FA} = \frac{c_{\alpha(2)}}{2K\pi}\left(\frac{18}{M}10^{-\mathrm{SNR}/10}+\frac{72}{M^2}10^{-2\mathrm{SNR}/10}\right) \tag{B.1}$$

B.2 Equivalence of two QPC detectors

The QPC detector described in Chapter 4 and [6] has similarities with one proposed about the same time by other researchers [5, 4].

In this Appendix it is demonstrated that, for the case of signals like $x(n)$ in Equation 3.4, with real sinusoids in AWGN (additive second-order white Gaussian noise), the two detectors are in fact the same. This is shown by reformulating the expression for $\sigma^2_{\hat{\theta}}$, the biphase variance at the discrete bifrequency which corresponds to (f_1, f_2) (Equation 4.7 in Section 4.5, and from [72]), in terms of the model parameters A_i $i = 0, ..P-1$ and the noise variance σ^2_v used in [5, 4].

For clarity the abbreviation Υ will be used to denote the detector developed in this thesis (and first published in [6]) and Ψ will be used to denote the other detector [5, 4]. The contrast between the approach taken by the two detectors can be seen in Figure B.2.

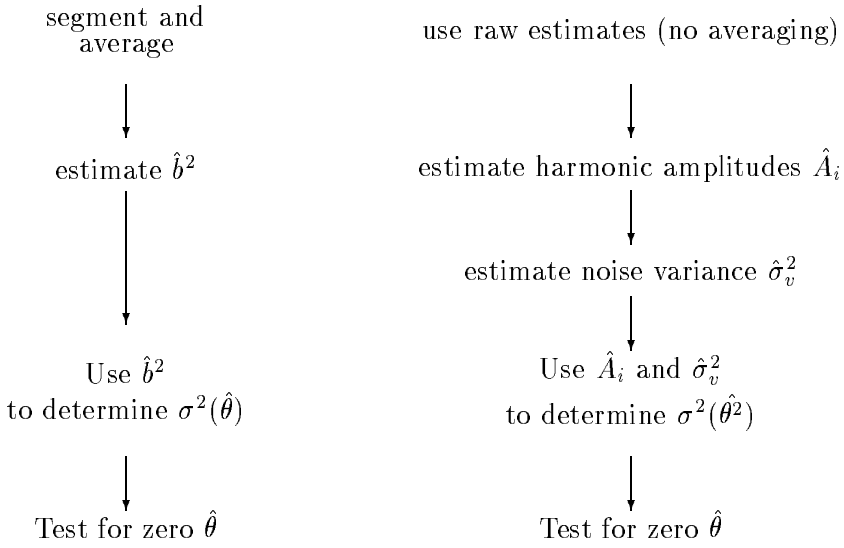

Figure B.1: Comparison between QPC detectors; Left : Υ detector Right : Ψ detector.

The use of segment averaging for the Υ detector should decrease estimate variances, and lead to improved statistical stability over the Ψ detector, which uses no averaging. It is thought that the reason why the Ψ detector does not use any averaging is because the results on which it is based are all *asymptotic*[5], and the view could be taken that because the estimates concerned are asymptotically unbiased and consistent then no averaging is required. However, from a more practical point of view, dealing with finite data lengths, the high variances of raw estimates can be a problem, and so some averaging is always a sensible option.

The key area in which the two detectors differ is in the expressions they use for the variance of the biphase estimate;

Υ The empirical variance of the biphase estimate at the discrete bifrequency corre-

sponding to (f_1, f_2) computed for a real signal over K independent frames [72];

$$\sigma^2(\hat{\theta})_\Upsilon \approx \frac{1}{2K}\left(\frac{1}{b^2_{f_1,f_2}} - 1\right), \tag{B.2}$$

Ψ The theoretical variance of the biphase estimate at the discrete bifrequency corresponding to (f_1, f_2) for a single record (of length M) of a signal consisting of complex harmonics in noise. This is the asymptotic variance of $\sqrt{M}(\hat{\theta} - \theta)$ [5, 4] divided by M,

$$\sigma^2(\hat{\theta})_\Psi \approx \frac{1}{2M}\sum_{i=1}^{3}\frac{\sigma_v^2}{A_i^2}. \tag{B.3}$$

Although the specific implementations of the detectors are different, it can be shown, in a variety of ways, that they are in fact very similar. This will be demonstrated below by reformulating the expression for the biphase variance from Equation B.3 for a signal composed of real sinusoids, estimated using segment averaging.

Consider the real signal model of Eqn. 3.4. For the purpose of this illustration it is assumed that the additive noise $v(n)$ is AWGN with variance σ_v^2, but what follows should be equally applicable to any second-order white noise with a symmetric pdf. Furthermore, for reasons of clarity it is also assumed that there are only 3 sinusoids, and their amplitudes are all the same $A_0 = A_1 = A_2 \equiv A$, although a similar, but slightly more complicated proof, can be carried out to show that the relationship is the same if this is not the case [88].

Equation 4.6 relates the squared bicoherence $b^2(f_1, f_2)$ (i.e. the squared bicoherence at a bifrequency corresponding to sinusoid frequencies f_1, f_2 and $f_1 + f_2$) to the sinusoid amplitudes A and the noise variance σ_v^2. Thus $\sigma^2(\hat{\theta})$ from Equation B.2 can be rewritten as

$$\sigma^2(\hat{\theta})_\Upsilon \approx \frac{1}{2K}\left(\frac{12}{M}\frac{\sigma_v^2}{A^2} + \frac{32}{M^2}\frac{\sigma_v^4}{A^4}\right) \tag{B.4}$$

Now from the asymptotic distribution of the biphase of a signal composed of *complex* harmonics $y(n) = \sum_i A_i \exp(-j2\pi f_i n + \theta_i)$ in *complex* noise [5, 4], an expression for *real* signals such as Eqn 3.4 can be formed. The results derived in [5, 4] are still valid, as

long as each instance of the squared sinusoid amplitude A_i^2 is scaled down by a factor of 4. This is because $E[|X_{f_i}|^2] = A_i^2$ for a complex harmonic signal[5, Equation 24], but $E[|X_{f_i}|^2] = A_i^2/4$ for a real signal (the total energy of the real signal is half that of the complex signal, and the real signal has half of its energy mirrored above the folding frequency). The biphase variance from Equation B.3 is then given by :

$$\sigma^2(\hat{\theta})_\Psi \Rightarrow \frac{1}{2M}\frac{3\sigma_v^2}{A^2/4} = \frac{12\sigma_v^2}{2MA^2}. \tag{B.5}$$

Furthermore, if K independent realisations are available, then the biphase variance is scaled down by a factor of K. The asymptotic biphase expression then leads to the following expression for the variance of the biphase averaged over K independent segments of $x(n)$;

$$\sigma^2(\hat{\theta})_\Psi = \frac{12\sigma_v^2}{2KMA^2}. \tag{B.6}$$

Under high SNR conditions the second bracketted term in Equation B.4 will be small. It is then apparent that under these conditions the expressions for biphase variance in Equations B.4 and B.6 reduce to the same form, and so $\sigma^2(\hat{\theta})_\Upsilon \approx \sigma^2(\hat{\theta})_\Psi$.

Appendix C

Experimental Method, Measures and Techniques

C.1 Speech Database specifications

In this section the composition of the speech database is described in terms of its design, size and specification, as well as describing how the recording sessions were conducted. The section also describes which of the recorded speech sounds were excluded from the bispectral analysis because of problems with the recorded speech.

C.1.1 The Speech Sounds

To build a database of speech sounds, it is difficult to describe to a subject, often with no knowledge of phonetics, which speech sound (e.g. "bilabial nasal") is required. A solution to this problem would be by imitation, i.e. to say to the speaker "please say the following sound in the way I say it", but this may not lead to the speaker using his/her *natural* speech sounds. A better solution is therefore to produce a list of words which contain the phonemes of interest. This is referred to as placing the phoneme *in context*.

A phoneme is an abstract unit [1], and so there is no guarantee that any two speakers will pronounce, for example "heed", in the same way. However, as long as subsequent analysis of the speech refers to *acoustic* rather than *phonemic* qualities of the speech sounds recorded, this will not matter at all.

Placing the sounds of interest in a word context does introduce the danger that coarticulation effects will be included in the database. However, careful segmentation and extraction of the "steady-state" part of the sound can minimise this danger, and the

placing of sounds in context makes the recording session significantly easier for the speakers.

The design of the test word list was thus composed of two parts

- Compile a list of the phonemes of interest (which are called the *target phonemes* here). These phonemes belong to the speech classes of vowels, fricatives (unvoiced and voiced) and nasals.
- Put each target phoneme in a suitable word context, trying to choose a context that has minimal influence on the target phoneme. A good way to do this for vowels is to put the vowel in a word beginning with "h" and ending with "t" or "d" [116], and this is possible for many of the words in the phoneme list. It is convenient to place the fricatives and nasals at the end of their context words, since speakers seem to find it easier to hold these sounds if they fall in these positions.

Table C.1 shows each class of speech sounds, the target phonemes (in both IPA [1] and SAM-PA [12] formats), and the phoneme-in-context test word. In each test word the underlined part indicates to the speaker the sound which is to be lengthened. The speakers were not given any instructions concerning the pitch of the sounds, and so each speaker chose a comfortable speaking pitch. Although no instructions were given to the speakers to produce a flat monotone pitch, it was found in the majority of cases that the speakers preferred such a delivery style. Consequently most of the utterances were spoken with a flat pitch.

C.1.2 The Speakers

The 16 speakers were volunteers from within the Departments of Electrical Engineering and Linguistics at the University of Edinburgh. Each speaker was given a 2-letter identification code, as shown in Table C.2 [1].

C.1.3 Recording Procedure

The recording equipment setup is described in Appendix C.2.

[1] These codes will be used to identify speakers in Chapter 7.

Sound Class	Target SAM-PA	Target IPA	Description	Test word
Fricatives (Unvoiced)	f	/f/	unvoiced labiodental	life
	S	/ʃ/	unvoiced palato-alveolar	wash
	s	/s/	unvoiced alveolar	bus
	T	/θ/	unvoiced dental	both
(Voiced)	v	/v/	voiced labiodental	love
	Z	/ʒ/	voiced palato-alveolar	leisure
	z	/z/	voiced alveolar	buzz
	D	/ð/	voiced dental	loathe
Vowels	i	/i/	high front	heat
	u	/u/	high back	hoot
	A	/ɑ/	low back	hart
	{	/a/	low front	hat
	V	/ɑ/	low central	hut
	I	/ɪ/	mid-high front	hit
	E	/ɛ/	mid-low front	head
	Q	/ɒ/	low back	hot
	@	/ə/	mid central	hurt
	O	/ɔ/	mid-low back	caught
	eI	/eɪ/	dipthong	hate
	U	/ɷ/	mid-high back	hood
Nasals	m	/m/	bilabial	ham
	n	/n/	alveloar	sin
	N	/ŋ/	velar	sing

Table C.1: Table showing target speech sounds in database with IPA and SAM-PA notations. Although all sounds recorded were analysed, most of the results presented in Chapter 7 consider only the boxed sounds.

Accent	Male	Female	Total
Southern English	*mc*, *jw*	*ca*	3
Northern English	*pb*, *gu*, *mb/md*	*cs*	4
Scottish	*is sw lb*	*kh*, *eb*,*rw*	6
Other	*jf* (N.Irish) *ps* (German)	*ta* (Greek)	3
Total	10	6	16

Table C.2: Composition of Speech Database

Each recording session was carried out in five sections, one each for vowels, fricatives (voiced and unvoiced) and nasals, as well as a section of digit reading (which was not subsequently used). As described above, each speaker was presented with a list of context words, an example of which is shown in Table C.3. In the list, the underlined section of each test word indicates the part of the sound to be lengthened.

First the speaker read aloud just a few of the words, to get used to the idea of lengthening a part of the word, and for the recording levels to be adjusted.

Then the speaker began reading aloud through the list, item by item. Some speakers experienced difficulties lengthening the sounds, and some records were repeated because of this.

index	word	OK	index	word	OK	index	word	OK	index	word	OK
01	h<u>ea</u>t		06	h<u>i</u>t		11	h<u>ea</u>d		16	h<u>a</u>t	
02	h<u>ar</u>t		07	h<u>oo</u>d		12	h<u>o</u>t		17	h<u>a</u>te	
03	h<u>u</u>t		08	h<u>ur</u>t		13	c<u>augh</u>t		18	h<u>oo</u>t	
04	h<u>ea</u>t		09	h<u>i</u>t		14	h<u>ea</u>d		19	h<u>a</u>t	
05	h<u>ar</u>t		10	h<u>oo</u>d		15	h<u>o</u>t		20	h<u>a</u>te	

Table C.3: Example of list of test words presented to speakers.

C.1.4 Subjective data checking

After the complete database had been assembled, every speech and laryngograph time series was plotted and inspected to check for;

1. Distortion arising from clipping.
2. Noise effects arising from low signal amplitude - resulting in evidence of quantisation noise, or poor SNR.
3. Misalignment of recording window with spoken utterance. For example sometimes the consonants surrounding the vowel of interest were present in the data file. In other cases the speech sound finished too soon.
4. Large amplitude variations.
5. Transient contamination (due to extraneous noises).

It was found that the record of one speaker *mb* had a very low amplitude, and so the complete record of that speaker was rerecorded (as *md*), and the original recordings discarded. Besides that a number of individual data files were excluded from further analysis. The files which were removed are listed in Table C.4.

sound type	speaker	SAM-PA	test word	file index	reason for exclusion
unvoiced fricatives	*ca*	f	life	13	transient
	is	T	both	16	amplitude variation
	kh	T	both	08	transient
	kh	T	both	12	transient
	mb	-	-	all	amplitude too low
voiced fricatives	*jf*	Z	leisure	06	clipping
	mb	-	-	all	amplitude too low
vowels	*mb*	-	-	all	amplitude too low
nasals	*mb*	-	-	all	amplitude too low

Table C.4: Files which were removed from analysis.

C.2 Equipment

Figure C.1 shows schematically how the different pieces of recording equipment were connected together. A small recording booth was constructed from five fabric-covered display panels, with a carpetted floor. The booth provided some sound insulation from outside noise sources, and had no significantly audible reverberation effects. An Audio-Technica ATM73a headset microphone was used to record the speech, and a Laryngograph[2] was used to record the laryngeal activity.

The laryngograph consists of two electrodes mounted in a neckband. The electrodes are placed on opposite sides of the larynx, and are connected to a control unit. The signal which comes out of the control unit, known as the "Lx signal", is very useful for detecting the instants of glottal closure. The laryngographic method of pitch extraction is generally more reliable than speech-based pitch extraction methods.

The signals from these two transducers were recorded directly to the two stereo channels of a Gravis PC soundcard mounted on a 486 PC, which was running the acquisition software "Clinassist[3]" The software is able to display the test words on the computer

[2] Manufactured by Larygograph Ltd, London, UK.

[3] written by Dr Alan Wrench of the Centre for Speech Technology, University of Edinburgh.

monitor, and simultaneously display the speech and Lx waveforms as they are recorded. The acquisition rate for both channels was 22.05 kHz. The gain on each channel was adjusted to prevent overload. For each channel this could be done in two ways, by adjusting an analogue potentiometer, or by moving a slider on the computer monitor.

As discussed in Section C.1.1 above, each sound of interest was placed in a word context, but the context itself was of no interest to the study, each recording was not triggered until the speaker had arrived at an approximately steady state of the sound of interest. For example, in the utterance "heed" the recording did not start until the speaker had reached the "ee" part.

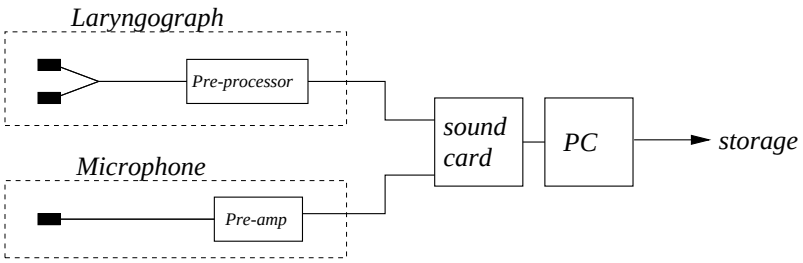

Figure C.1: Schematic diagram of recording set-up.

C.2.1 Microphone Phase

The microphone used in the data acquisition was an Audio Technica ATM73 headset microphone[117] with preamplifier. As it has already been discussed in Section 4.5.4.1 it is important that any transducers used in HOS analysis have linear phase characteristics, since such filters do not effect the QPC-detecting properties of the bicoherence. The measurement of the *actual* phase characteristic of any microphone is a difficult operation, requiring specialised equipment. No technical data concerning the phase response of this microphone was available from [117]. However, the linearity of the phase can be checked using fairly simple equipment.

Figure C.2 shows the experimental setup for a phase calibration procedure that took place at the Institute of Sound and Vibration Research, University of Southampton. The sound source is a hi-fi loudspeaker, driven by white Gaussian noise. The test microphone, and a reference microphone with known phase response, are placed close to each other, about 10cm from the source. The signals from each microphone are then fed into a two-channel DFT analyser, where the transfer function can be calculated. Under certain assumptions, which will be mentioned below, the sound field at each

microphone can be assumed to be the same. The complex cross-spectrum between the two microphone measurements $P_{xy}(k)$ then provides information about the test microphone phase response.

The phase of the cross spectrum is related to the phase responses of the microphones by

$$\phi(k) = \phi_{\text{ref}}(k) - \phi_{\text{test}}(k) \tag{C.1}$$

Now in this experiment the reference microphone is a Brüel and Kjær 4136 1/4 inch condenser microphone. From data books [118, p71] it is known that this microphone has a linear phase characteristic

$$\phi_{\text{ref}}(k) = -\alpha k. \tag{C.2}$$

Consequently, if $\phi_{\text{test}}(k)$ is also linear, the phase of the cross spectrum will be

$$\begin{aligned} \phi(k) &= -\alpha k - \beta k & \text{(C.3)} \\ &= -(\alpha + \beta)k. & \text{(C.4)} \end{aligned}$$

i.e. *If the cross spectral phase is linear, then this implies that the phase response of the test microphone is also linear.*

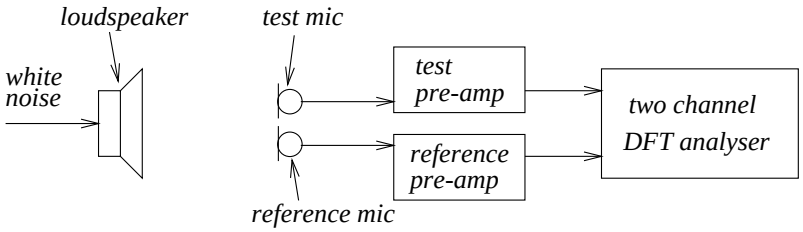

Figure C.2: Experimental set-up for microphone phase calibration.

The assumption that both microphones are in the same sound field is only valid so long as the microphone separation is significantly smaller than the smallest sound wavelength (i.e. the wavelength of the highest frequency sound). Given that the microphone separation was $\approx$ 5mm, it can be assumed that this method is reliable down to wavelengths $\approx$ 5cm, which corresponds to a frequency of $f = c/\lambda = 330/.05 \rightarrow 6.6$ kHz.

Figure C.3 shows the phase of the cross-spectra between the two microphones. It is evident that, up to the higher frequency limit of 6.6 kHz established above, the phase

has an approximately linear characteristic.

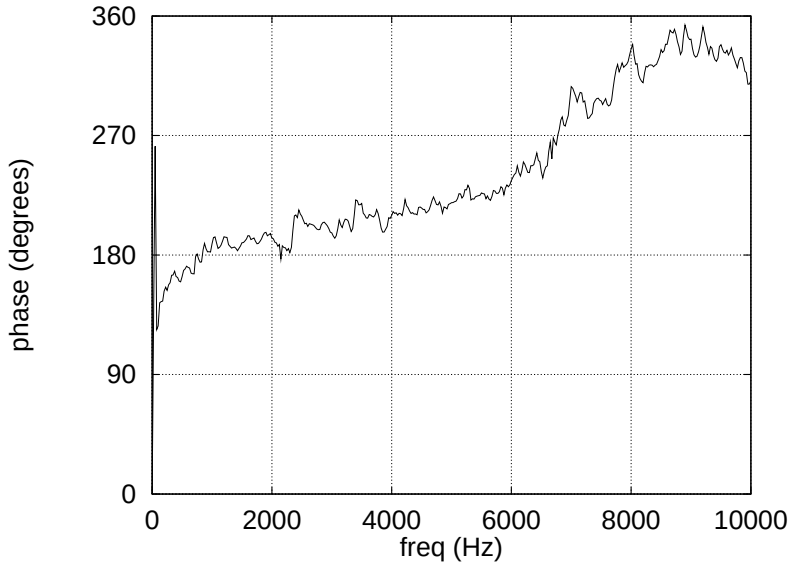

Figure C.3: Phase of cross-spectrum between reference and test microphones.

C.2.2 Storage requirements

The data was stored as 16-bit binary, with each speech sound lasting $1.81s$ (40000 samples at 22.05kHz). The memory requirements S for the database depend on the following quantities (with the actual values used written in braces);

N_{spkrs} the number of speakers {16},

$N_{samples}$ the number of samples in each record {40000},

N_{mic} the number of speech sounds for which microphone records are needed {23},

N_{lx} the number of speech sounds for which laryngograph records are needed (the voiced sounds only) {19},

N_{utt} the number of repetitions of each sound {5}.

b the number of bits per stored number {16}.

The storage requirement is then calculated from

$$S \approx N_{spkrs} \times \frac{b}{8} \times N_{samples} \times N_{utt} \times [N_{mic} + N_{lx}] \quad \text{(C.5)}$$

$$= 269\text{Mbytes} \quad \text{(C.6)}$$

It is clear that the data storage requirements are quite heavy.

C.3 Spectral Moments

In this section the spectral moments are defined, following the method described elsewhere [106].

The key to the method is to treat the normalised power spectrum as though it was a probability distribution. The power spectrum $P(k)$ is first normalised

$$p'(k) = P(k) / \sum_{l=0}^{l=M} P(l). \tag{C.7}$$

The first spectral moment is defined as

$$m_1 = \sum_{l=0}^{l=M} l p'(l), \tag{C.8}$$

and the second spectral moment as

$$m_2 = \sum_{l=0}^{l=M} [l - m_1]^2 p'(l) \tag{C.9}$$

The third and fourth spectral moments are computed in a similar way, but with an additional normalisation

$$m_3 = \sum_{l=0}^{l=M} [l - m_1]^3 p'(l) / m_2^{3/2} \tag{C.10}$$

$$m_4 = \sum_{l=0}^{l=M} [l - m_1]^4 p'(l) / m_2^2 - 3 \tag{C.11}$$

Although not based on rigorous mathematics, the spectral moments are in widespread use as tools to characterise the spectra of speech sounds, in particular fricatives. It is important to note that the spectral moments have no meaningful relation to the HOS moments, or polyspectra. The assumptions underlying computation of spectral moments are

- The signal is stationary.
- The power spectrum, when viewed as a pdf, is a uni-modal distribution (i.e. it contains one main peak only). This is a gross approximation since it is widely understood that all speech sounds, even fricatives, have spectra with several peaks.

Spectral moments are usually computed from single-shot power spectra [106], but there is good reason to obtain better estimates using some form of averaging [119]. In previous work [120] ensemble averaging was used to obtain reliable spectral estimates, but since in the current study extended durations of each utterance were available, a segment-averaging approach was possible.

C.4 Pitch-Synchronous (PS) Analysis of Voiced Speech

Chapter 6 identifies the motivation for carrying out bispectral analysis pitch-synchronously. The technique involves the extraction of a fixed number of samples (determined by the DFT size M) from each pitch cycle.

To carry out voiced speech analysis *pitch-synchronously* a pitch reference signal is needed. This reference signal can either be computed directly from the speech signal [103] or measured using a laryngograph[4].

For voiced speech the event of most importance as far as vocal tract excitation is concerned is that of the glottal closure. As discussed in Chapter 2 the glottis snaps shut rapidly, and this constitutes the maximal excitation of the vocal tract. Although it is sometimes difficult to identify the point of glottal closure from the speech signal, it is straightforward to identify it from the Lx signal, since it corresponds to the time at which the Lx signal is changing most quickly.

C.4.1 Basic Algorithm

Figure C.4 shows an example of the speech sound `i` (as in "heed") at various stages of procesing. The main stages are as follows;

- Differentiate the Lx signal $l(n) \quad n = 0, .., N-1$ to form a new record $l'(n) \quad n = 1, .., N-1$;

$$l'(n) = l(n) - l(n-1) \tag{C.12}$$

[4] See Appendix C.2 for details of the equipment used.

- To extract the glottal closures from $l'(n)$, apply a threshold to $l'(n)$, resulting in an impulse train $i(n)$, with impulses ('1's) corresponding to glottal closures. It has been found that a threshold setting of $\frac{1}{5}\max l'(n)$ gives good performance.

  In order to prevent the problem of detecting several closures as the result of one actual closure, a simple rule is used that prohibits the detection of any new closure until a specified time has passed since the last detected closure. This time is set to f_s/f_{max} where f_{max} is the maximum permittable pitch frequency, 200 Hz.

- The processed signal is now a pulse train;

$$i(n) = \sum_{m=0}^{m=N_p-1} \delta_{z(m)} \quad (n = 1, .., N-1) \tag{C.13}$$

  where $z(0), z(1), .., z(N_p - 1)$ are the indices (sample numbers) of the glottal closures, and δ is the Kronecker delta function, which is 1 if $n = z(m)$, and 0 otherwise. There are N_p pulses in this time frame.

- Apply a delay Δ to the $z()$ to account for the time difference between the Lx measurement and the delayed acoustic signal $s(n)$. This delay will depend on the propagation distance between larynx and microphone and the speed of sound in air. Assuming the average vocal tract to be of length 15cm and the average speed of sound in (humid) air to be 350ms^{-1} this gives a delay of ≈ 0.5ms [13], or 10 samples with $f_s = 22.05$ kHz[5]

- Extract a window of M speech samples from $s(n)$ for each pulse in $i(n)$, discarding the part of the pitch period outside this window.

C.4.2 Choosing the window centre position

In its simple form, the algorithm described above will extract speech frames in which the glottal excitation is at the extreme left- hand-side (LHS) of the window. Whilst this is a good state of affairs from a physical point of view, since causality dictates that the response of the vocal tract to the event of glottal closure event will occur at a later time

[5] Of course different assumptions concerning the VT length and speed of sound could result in different delays, but changing the parameters will only change the delay by 2 or 3 samples, and so the choice of these parameters is not considered to be a very important component in the analysis.

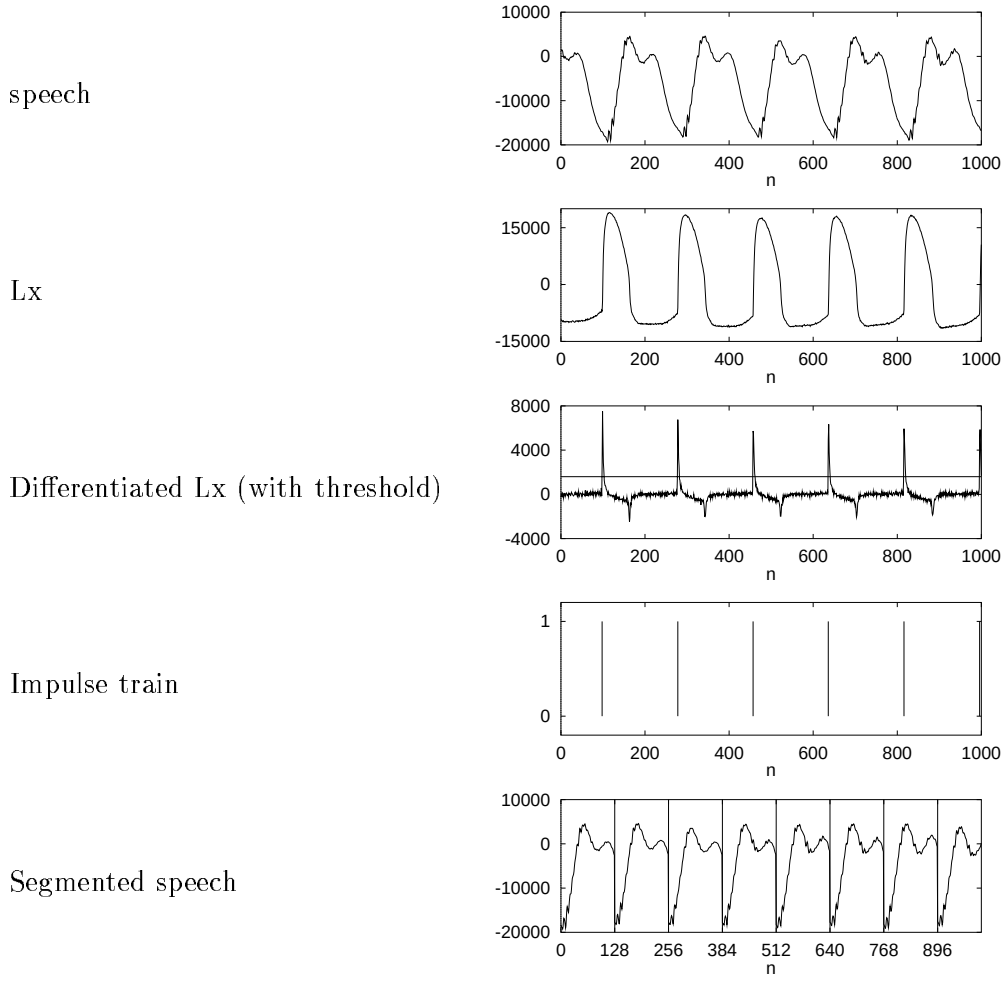

Figure C.4: A vowel sound in various stages of pre-processing for PS analysis. In this example $M = 128$.

than the glottal closure itself, it may be less than ideal from a signal processing point of view, for the following reason; waveform information occurring close to the instant of glottal closure may be severely reduced if a "soft" data window (e.g. Hamming) is used.

Although a boxcar (rectangular window) preserves all waveform information in the data frame, it results in the undesirable spectral [84] (and bispectral [71]) property of leakage. Soft windows have better leakage properties, but they do this by weighting down the waveform at the frame edges.

A schematic representation of this problem is shown in Figure C.5. This shows three concatenated pitch cycles extracted using the PS algorithm described above. It can be seen that the speech samples corresponding to the moment of glottal closure will be strongly reduced.

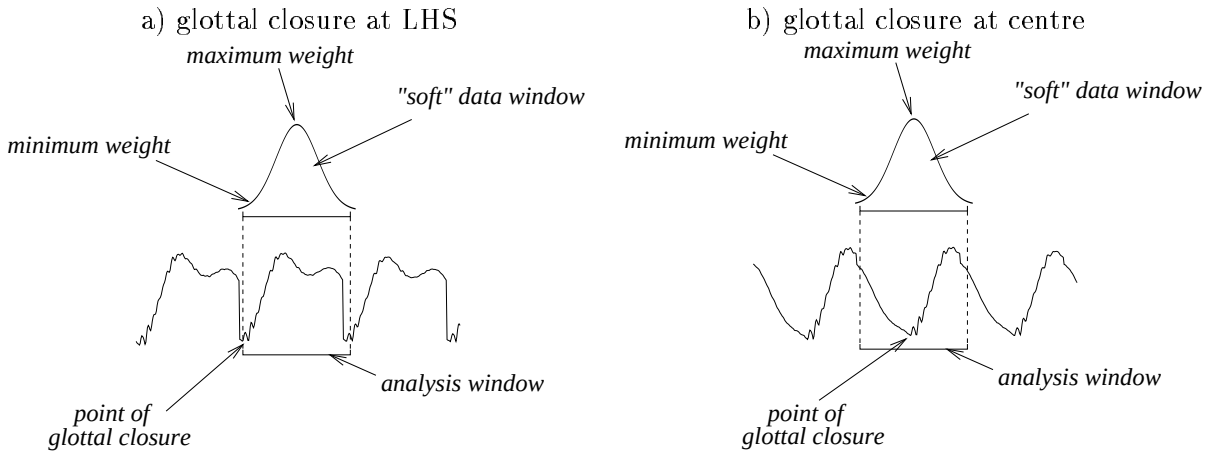

Figure C.5: Example of 2 different approaches to windowing; a) Glottal closure at LHS of analysis window b) Glottal closure at middle of analysis window.

A solution to this problem can be found by sliding the glottal closure point by some as yet unkown number of samples to the right. This means that the instant of glottal closure is moved towards the centre of the analysis frame, and it consequently receives a greater weight after windowing, as shown in Figure C.5 b).

The amount by which the glottal closure instant is moved in the analysis window is subject to the following constraints;

physics The part of the waveform which is analysed should be focussed on the instant of glottal closure, and samples immediately following it.

signal processing The instant of glottal closure should not be too close to the analysis window boundaries, because otherwise it will be weighted down too much if soft data windows are used.

Figure C.6 shows the time series (after segmentation) and power spectra of the sound `i` processed with the glottal closure at three different positions in the analysis window. It is evident that if the glottal closure is placed at the beginning of the frame (part a) of Figure C.6) , there is signal energy at many frequencies. As the glottal closure is moved towards the centre of the frame (figures b) and c)) the high frequency energy is reduced. This is thought to be due partly to the fact that waveform information is lost at the right hand edge of the analysis window as the glottal closure instant is moved towards the frame centre.

Using the spectral magnitude as a rough measure of the amount of information encapsulated in the analysis frames, it appears that the advantage of moving the glottal closure point to the right is outweighed by the loss of information at the right-hand edge of the frame.

Consequently, in the analyses presented in the thesis, the glottal closure point is fixed at the LH edge of the analysis window (as in part a) of Figure C.6, but this may be an area in which more research is warranted. In particular it may be of interest to develop bispectral techniques which use asymmetrical data windows which are designed specifically for pitch-synchronous analysis. Such windows would give most weight to the 'early' part of the waveform (i.e. just after the glottal closure) and then decay exponentially in time.

C.4.3 Length of data required for PS Analysis

As mentioned in Section 6.4.1, the data length requirement for bispectral estimates of unvoiced speech needs some modification for PS analysis of voiced speech. For voiced speech, with averaging over K segments, K pitch periods are required. This means that the data record must be of sufficient length T (in s) to satisfy

$$T \geq K/f_0 \tag{C.14}$$

where f_0 is the speech fundamental frequency (assumed approximately constant for this calculation). Given that f_0 is typically somewhere between 100 Hz (male) and 200 Hz

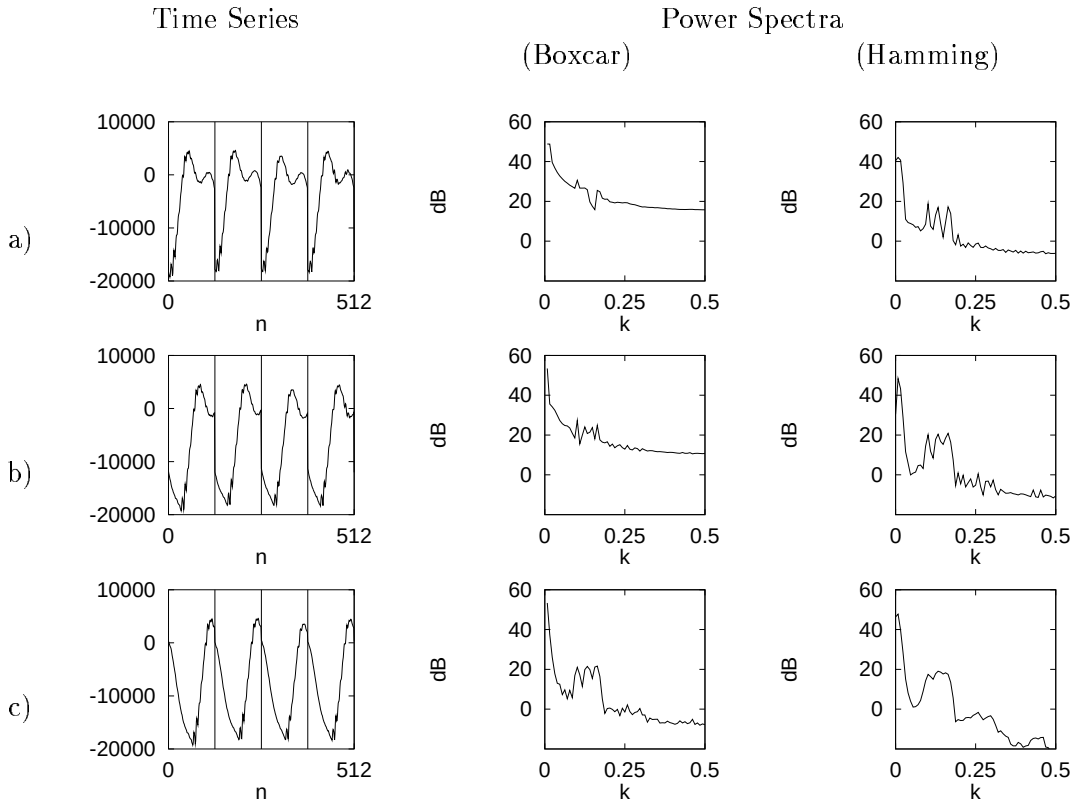

Figure C.6: Time series (left) and Power spectra (right) of speech sound `i` with glottal closure placed at different positions in analysis window; a) at LHS of window, b) 1/4 of a window's width from LHS c) at middle of window. $f_s = 22.05$kHz, $N = 4096$, $M = 128$ (so frame durations are same as for $M = 64$ at $f_s = 11.025$kHz). Frequencies shown on normalised scale f/f_s.

(female), and assuming that 64 data segments are required, then this means that in the worst case (when the voicing frequency is the lowest) the length of data record must be at least $T \geq 64/100 = 0.64$s, i.e. that for male speakers, at least 0.64s of steady-state speech are required to carry out a 64-point DFT bicoherence analysis. Note that, in contrast to the data length requirement for unvoiced speech, the requirement for voiced speech (Equation C.14) does not depend on the sampling rate.

Appendix D
Statistical Analysis

D.1 Levene Test

Before embarking on an ANOVA (analysis of variance) investigation, it has to be verified that the statistical properties of the data match the requirements of the ANOVA method. Central to these requirements is that the variance of the data is homogeneous [109, 110].

The Levene test for homogeneity of variance is attractive because it is not particularly dependent on the assumption of normality in the sampled data[109, p66]. It involves a one way analysis of variance on the absolute values of the residuals. An example of its application to the type of data considered in this thesis will be presented here for illustrative purposes.

Consider a subset of the results data for one measurement variable (say X) only. Now consider the sample values of that variable for *one sound only*, with all $N_{utt} = 5$ utterances, and for all $N_{spkrs} = 16$ speakers. The total number of utterances, N_{Σ}, is then $16 \times 5 = 80$.

That is the data is of the form X_{ij} where i is the speaker number and j is the utterance number.

	utterances	
speakers	X	Mean
speaker 1	X_{11}, X_{12}, X_{13}, X_{14}, X_{15}	$\overline{X}_1$
speaker 2	X_{21}, X_{22}, X_{23}, X_{24}, X_{25}	$\overline{X}_2$
...	.....	..

Now the matter under investigation is how the variances of the samples from different speakers are related. This is done by carrying out a one-way analysis of variance on

the residuals $X_{ij} - \overline{X}_i$, where $\overline{X}_i$ is the average of the measurements on speaker i. Under the null hypothesis H_0, that the variances are equal[1], the Levene statistic is F-distributed with $N_{spkrs} - 1$, $N_\Sigma - N_{spkrs}$ degrees of freedom. If the computed statistic is greater than the critical value c for $F(N_{spkrs} - 1, N_\Sigma - N_{spkrs})$ at the 5% significance level (for example), then H_0 is rejected.

It is known [109] that care is required when applying this test to small samples.

D.2 Homogeneity of Variances

Tables D.1, D.2 and D.3 show the results of the Levene tests for homogeneity of variance on the fricatives, vowels and nasals respectively, for the feature $\sum_{\mathrm{IT}} b^2$ (introduced in Section 6.4.2). In each table entry, the null hypothesis H_0 being tested is the hypothesis that the variance of $\sum_{\mathrm{IT}} b^2$ across the multiple utterances (usually 5) by speaker 1 is the same as the variance across multiple utterances by speaker 2, and so on for all speakers.

If the F-statistic shown in the table is greater than the threshold for the required number of dof, then H_0 is rejected, and the alternative hypothesis - that the variances are not homogeneous - is accepted. These statistics are shown in **bold**.

Statistics for which H_0 is accepted are shown in *italic*. It is evident in Table D.1 that some phonemes `f` and `Z` have only 15,63 dof, whereas all other sounds have 15,64 dof. This is because a few files were excluded from further analysis in the data preparation stage described in Appendix C.1.4.

The results for Tables D.1, D.2 and D.3 indicate that, for the majority of test phonemes, the variance is *not* homogeneous, and so ANOVA analysis of the data in its current form may be misleading. In practical ANOVA analysis, it is often possible to carry out a tranformation on the data to ensure homogeneity of variance, but that line of investigation has not been pursued any further in this thesis.

[1] i.e. that the variance of the measurements of X for speaker 1 is the same as the variance of the measurements for speaker 2 etc.

			Levene Test	
			Conventional estimator	Robust estimator
SAM-PA	dof		F-statistic	F-statistic
f	15	63	**2.39**	**2.02**
S	15	64	**2.08**	*0.97*
s	15	64	**2.03**	*0.50*
T	-	-	-	-
v	15	64	**4.08**	**9.42**
Z	15	63	**4.57**	**7.83**
z	15	64	**17.10**	**17.11**
D	15	64	**4.60**	**9.03**

Table D.1: Results of Levene tests for homogeneity of variances between speakers of the feature $\sum_{\mathrm{IT}} b^2$ for each word. For a significance level of 0.05, F-statistics shown in **bold** reject H_0 and those shown in *italics* accept H_0. No statistics could be calculated for T because the number of utterances available for speaker *kh* was only 3.

			Levene Test	
			Conventional estimator	Robust estimator
SAM-PA	dof		F-statistic	F-statistic
i	15	64	**2.21**	**2.05**
u	15	64	**1.89**	**2.73**
A	15	64	**1.88**	**2.22**
{	15	64	*0.96*	*1.10*
V	15	64	**2.37**	**2.21**
I	15	64	*1.31*	*1.14*
E	15	64	*1.27*	*1.50*
Q	15	64	**4.38**	**5.47**
@	15	64	*1.21*	*1.11*
O	15	64	**2.30**	**2.35**
eI	15	64	*1.55*	*1.23*
U	15	64	*1.78*	**3.58**

Table D.2: Results of Levene tests for homogeneity of variances between speakers of the feature $\sum_{\mathrm{IT}} b^2$ for each word. For a significance level of 0.05, F-statistics shown in **bold** reject H_0 and those shown in *italics* accept H_0.

			Levene Test	
			Conventional estimator	Robust estimator
SAM-PA	dof		F-statistic	F-statistic
m	15	64	**2.25**	**3.36**
n	15	64	**10.93**	**12.00**
N	15	64	**3.56**	**4.19**

Table D.3: Results of Levene tests for homogeneity of variances between speakers of the feature $\sum_{\mathrm{IT}} b^2$ for each word. For a significance level of 0.05, F-statistics shown in **bold** reject H_0 and those shown in *italics* accept H_0.

Appendix E
Publications

J.W.A. Fackrell and S. McLaughlin, "The Higher Order Statistics of Speech Signals," *IEE Colloquium on Techniques in Speech Signal Processing*, London, pp 7/1-7/6, 1994.

J.W.A. Fackrell and S. McLaughlin, "Detecting Phase Coupling in Speech Signals," *IEE Colloquium on Speech and Image Processing*, London, pp4/1-4/8, 1995.

J.W.A. Fackrell and S. McLaughlin, "Detecting Phase Coupling Using the Bicoherence," *IEE Colloquium on Higher Order Statistics*, London, pp9/1-9/8, 1995.

J.W.A. Fackrell, S. McLaughlin and P. R. White, "Practical Issues Concerning the Use of the Bicoherence for the Detection of Quadratic Phase Coupling", *IEEE Signal Processing Workshop on Higher-Order Statistics*, Spain, pp 310-314, 1995.

J.W.A. Fackrell, S. McLaughlin and P.R. White, "Bicoherence Estimation using the Direct Method : Practical Issues", accepted for publication in two parts in *Applied Signal Processing*, 1996.

J.W.A. Fackrell, A.G. Stogioglou and S. McLaughlin, "Robust Frequency-Domain Bicoherence Estimation", *8th IEEE Signal Processing Workshop on Statistical Signal and Array Processing*, Corfu, pp 206-209, 1996.

J.W.A. Fackrell and S. McLaughlin, "Determining the False-Alarm Probability of HOS-based Quadratic Phase Coupling Detectors", *Proceedings of EUSIPCO 96*, Trieste, Italy, 1996.

J.W.A. Fackrell and S. Mclaughlin, "Detecting nonlinearities in speech sounds using the bicoherence", *Proceedings of the Institute of Acoustics Autumn Conference, Speech and Hearing 96*, Windemere, Vol 18, Part 9, pp 123-130, 1996.

J.W.A. Fackrell and S. Mclaughlin, "Robust Non-parametric Bicoherence Estimation by Stepwise Outlier Rejection", submitted to IEEE Signal Processing Letters, 1996.