The Application of Auditory Signal Processing Principles to the Detection, Tracking and Association of Tonal Components in Sonar

Robert William Mill

Department of Computer Science
University of Sheffield

November 2008

Dissertation submitted to the University of Sheffield
for the degree of Doctor of Philosophy

Abstract

A steady signal exerts two complementary effects on a noisy acoustic environment: one is to *add energy*, the other is to *create order*. The ear has evolved mechanisms to detect both effects and encodes the fine temporal detail of a stimulus in sequences of auditory nerve discharges. Taking inspiration from these ideas, this thesis investigates the use of regular timing for sonar signal detection. Algorithms that operate on the temporal structure of a received signal are developed for the detection of merchant vessels. These ideas are explored by reappraising three areas traditionally associated with power-based detection.

First of all, a time-frequency display based on timing instead of power is developed. Rather than inquiring of the display, "How much energy has been measured at this frequency?", one would ask, "How structured is the signal at this frequency? Is this consistent with a target?" The auditory-motivated *zero crossings with peak amplitudes* (ZCPA) algorithm forms the starting-point for this study.

Next, matters related to quantitative system performance analysis are addressed, such as *how often* a system will fail to detect a signal in particular conditions, or *how much* energy is required to guarantee a certain probability of detection. A suite of optimal temporal receivers is designed and is subsequently evaluated using the same kinds of synthetic signal used to assess power-based systems: Gaussian processes and sinusoids.

The final area of work considers how discrete components on a sonar signal display, such as tonals and transients, can be identified and organised according to auditory scene analysis principles. Two algorithms are presented and evaluated using synthetic signals: one is designed to track a tonal through transient events, and the other attempts to identify groups of comodulated tonals against a noise background. A demonstration of each algorithm is provided for recorded sonar signals.

Acknowledgements

I would firstly like to thank Guy Brown both for his careful supervision of my work over the past four years and his illimitable encouragement. "Time is moving on, but it's not running out." My thanks also go to the other two members of my thesis panel, Jon Barker and Joab Winkler, for their insightful comments throughout the degree. They have kept their office doors open for me to ask questions, seek advice or borrow stuff at any time.

I would like to express my gratitute to Andrew McLean and Patrick Tindell of the Marine and Acoustics Department at QinetiQ for their financial and technical support. Specific thanks are due for supplying the recorded sonar data used in this project, and for the ideas shared and hospitality extended at our meetings in Winfrith.

Oded Ghitza (Sensimetrics Corporation) deserves a quick note of appreciation for the useful e-mail correspondance shared concerning his auditory model.

Gracing the dusty corridors of academia once again... I am grateful to Mike Stannett for his input concerning the material in the fifth chapter. Very warm thanks are due to Phil Green, Richard Clayton, Mahesan Niranjan and Martin Cooke for giving me so many fantastic opportunities to teach. I would like to thank all the members of the Speech and Hearing Research group for collectively crafting such a friendly environment to work in. I particularly thank James, Jonny and Stu, for their friendship and interest in my work, along with Simon T.: a veritable lending library for arcane journal papers and odd music.

I would especially like to thank my parents for the constant love and encouragement they have shown me throughout my years of formal education and the four years of informal education that went before them! I must also thank Dave and John for their examples of a Christian life pursued in science, and my grandmother for all the lengthy phone calls. A big thankyou to Els's parents too, for their financial support and kindness.

Credit goes to my trio of friendgineers, Martin, Pete and Rich, for their friendship and advice during my time in Sheffield. In time-honoured fashion,

```
10 PRINT "Thanks."
20 GOTO 10
```

I am indebted to Al and Chris, and James, and a countless crowd cheering me on, for supplying me and my wife with much-needed food and blankets during the cold Ph.D. winters. As the first in this crowd, it is my greatest pleasure to thank Jesus Christ, the Incarnate Word of God, who straightens my paths.

Last of all, I thank Els for her love, patience, hard work, understanding, patience and moral support. I dedicate this thesis to you.

Table of Contents

Chapter

1

Introduction

Imagine you are asked to give a non-technical description of the ear. How you would reply? You might respond that the ear turns sound waves into something that we can hear and understand. Alternatively, you could draw an analogy from the world of technology: microphones, telephones, audio surveillance devices and intercoms are all *mechanical sound-receivers*, and most people today would be content to view the ear (and perhaps other aspects of audition) as a biological mechanism for receiving sound. The starting point of this thesis is one such analogy: specifically, the analogy between hearing and *passive sonar*—a technology for receiving and analysing underwater sound waves.

This introductory chapter falls into three parts. First, the analogy between hearing and sonar must be scrutinised more closely. Does it emerge naturally from the empirical results of two, largely-separate sciences, or is it being over-eagerly applied on the basis of similarities that appear on the surface? The ear and passive sonar must also exhibit some differences as well as similarities, if aspects of the former are to inspire useful changes to the design of the latter. One such difference relates to how the temporal information in an acoustic signal is used.

The central argument of this thesis is that the auditory system utilises temporal features of a signal encoded in the timing of auditory nerve spikes to aid signal analysis, whereas a conventional sonar system relies exclusively on power; and, furthermore, that the success of the ear is partially due to this use of temporal information and thus motivates the investigation of a sonar receiver built on similar principles. The second section reviews various mathematical interpretations of the timing information in signals (specifically, zero crossings), with a view to considering which frameworks might be useful in later chapters. The third section provides an overview of the thesis objectives and indicates how the remaining material is to be structured.

1.1 The Groundwork for an Analogy

We will shortly examine how a biological receiver—the ear—has already informed, and has the potential to inform further, the design of a passive sonar receiver. However, it is prudent to begin by exploring the historical basis for the ear-and-sonar-as-receiver analogy, given its foundational role in what is to follow.

1.1.1 The Ear as a Receiver: Historical Perspectives

The mechanicity and recipience of the ear is taken for granted today, but it has not always been so. Pythagorus (575–500 B.C.) proposed that ears (and eyes) operate rather like the sense of touch, manipulating the environment directly from afar. Aristotle (384–322 B.C.), who sought to integrate his understanding of the ear with the prevalent classical theory of the four elements—air, fire, earth and water—concluded that the ear was the "organ of the air". So influential was his theory, that the inner ear was still referred to as *aer internus* [Ltn. internal air] by some workers until the late 18th Century.

Volcher Coiter (1534–1600), towards the close of the 16th Century, published *De auditus instrumento* [Ltn. *The Instrument of Audition*], which summarised the work of the earlier anatomists, notably, Vesalius, Fallopius and Eustachio (Finger, 2001). This account, as well as naming several parts of the ear, described how sound was collected by the outer ear, amplified by the middle ear and set the eardrum in motion. It also speculated (correctly) that the nerve endings in the cochlea were responsible for detecting sound, although a role was reserved for the *aer internus*—a vestige of classical natural philosophy. A more detailed anatomical picture emerged over the next two centuries, spurred on by the advent of the microscope and more delicate techniques for preparing specimens for dissection (von Békésy and Rosenblith, 1948).

An explanation of the *mechanics* of the ear was lacking in all but qualitative terms until relatively recently. Hippocrates (ca. 460–377 B.C.) and his contemporaries had observed a hollow cavity within the ear and duly suggested that hearing was somehow mediated by echoes or resonance (de Cheveigné, 2005; von Békésy and Rosenblith, 1948). But it was not until the time of Helmholtz (1821–1894) that a sufficiently-advanced mathematics of acoustics was available to explain quantitatively how parts of the ear resonated at different frequencies and could thereby decompose a sound—a concept which is still foundational in modern hearing science today (Moore, 2004).

Theories of hearing based on resonance were accepted slowly for various reasons. In the former half of the 19th Century, the empirical evidence was not compelling enough to settle the matter. As late as 1866, Rutherford (1839–1899), inspired by the latest advances in communications technology, proposed the *telephone theory*, which states that the ear transmits a copy of the stimulating waveform to the brain via impulses along the auditory nerve (Finger, 2001). This theory, in its original form, is now discredited; but a weaker version of the idea persists in the form of *temporal coding theories*, which allow that the auditory neural signal convey the vibration of resonating structures.

The abandonment of the telephone theory did not subdue comparisons between the ear and a telephone. In 1967, Zwicker and Feldtkeller published *Das Ohr als Nachrichtenempfänger* [Ger. *The Ear as a Communication Receiver*], with the following statement in the English translation of its preface (Zwicker and Feldtkeller, 1967/1999):

> "These questions [about the ear] are relevant to communication technology, because telephones and radios serve to transmit speech and music. [...] We would like to provide an understanding of how the auditory system performs as a receiver and as a measurement device."

The invention of analogue and digital hearing aids in the 20th Century undoubtedly did much to cement the notion of the ear as a mechanical receiver. The earliest devices selectively amplified a signal; later models used multiband signal processing. In the past thirty years, controlled stimulation of the auditory nerve via a cochlear implant has allowed partial restoration of hearing to those suffering from severe hearing loss (Moore, 2004).

With the ear being variously referred to down the ages as an "instrument", "telephone", "communication receiver" and "effective signal processor" (Dau et al., 1996), the analogy between the ear and mechanical receiver appears to be a well-founded one. We must now proceed to the second subject of the receiver analogy: sonar.

1.1.2 The Development of Sonar

Early underwater acoustics experiments in the 1800s involved striking a submerged bell and measuring the time taken for the sound to arrive at a remote location by signalling with lights. Over the course of the century, the propagation of sound in the sea was exploited for communication, warning systems and the passive detection of ships. The receivers took the form of tubes or "trumpets", one end of which was placed in the water, the other at the ear.

The discovery of the electrical generation of sound towards the end of the 19th Century naturally led to advances in underwater science and technology. In 1912, Fessenden (1866–1932) developed a single-frequency, electrical oscillator to replace the bell with a controlled, high-power source. The sinking of the *Titanic* ten years later prompted Fessenden to incorporate his electrical source into an *echo-ranging system* for detecting icebergs, a concept which still remains in modern-day active sonars.

The eruption of World War I (1914–1918) created an urgent need for new acoustic technologies for use in warfare. The American SC tube employed a stethoscope-like device consisting of two listening tubes, which ran down a central column into the water and turned in opposite directions at their ends. The listener would place one tube in each ear and then rotate the column to discover the azimuthal bearing of a sound. Several elaborations of the SC device followed shortly, including the addition of multiple receivers and the towing of receivers at a distance to reduce self-noise.

Technological progress in underwater acoustics slowed considerably between the world wars but was revived by the outbreak of World War II (1939–1945) and continued into the postwar era. Advances were made on a number of fronts: surveys of ocean sound speed and noise characteristics were undertaken in different locations around the world during different seasons, and major theoretical results in statistical signal processing and information theory were published (including those of Rice, Wiener, Gabor and Shannon). It was shortly after the war that the acronym *sonar*—an abbreviation of **so**und, **na**vigation and **r**anging—entered the international English vocabulary.

The advent of electronics and digital signal processing led to radical improvements in sonar technology. *Active sonar*, which investigates the underwater environment by transmitting a transient pulse and analysing the returning echoes (like Fessenden's early echo-ranging device), was gradually abandoned during the Cold War in favour of silent, *passive sonar*, which listens for the sounds radiated by targets themselves (like the SC tube). The need to detect the tonal emissions of quiet submarines over long distances thus grew in importance, and the invention of the fast Fourier transform (FFT) (Cooley and Tukey, 1965)—a computationally-inexpensive form of the discrete Fourier transform (DFT)—made narrowband processing a practical possibility. In modern-day research, underwater sound is routinely recorded by a hydrophone array, digitised, copied, transported and post-processed using a computer. For a more detailed history of underwater acoustics, the reader is directed to the treatments of Lasky (1977) and Burdic (1984, Chapter 1).

1.1.3 Points of Contact

By now we have hopefully established that passive sonar and the ear are analogues inasmuch as they are both mechanical sound receivers. To rule out a merely superficial comparison, we shall attempt in this section to establish multiple points of contact between the two subjects—particular aspects in which sonar and the ear are alike. It bears emphasising that the similarities revealed in these very short case studies are, to the knowledge of the author, *accidental*; that is, there was no conscious attempt on the part of the sonar engineers to mimic aspects of audition. In some cases, we will take the opportunity to give passing mention to areas of common ground that have not been explored in the thesis.

Passive and Active Listening

For a system to be described as either a sonar or a listener, we require: a vibrating source, a fluid medium through which the vibrations can propagate, a transducer and a classifier. Sound cannot exist without the first and second of these; neither hearing nor sonar can exist without all four. A *transducer* is any device that converts the sound energy into another form, e.g., mechanical, electrical or chemical. A *classifier*, for our purposes, is a component that extracts useful information about the environment from the transducer output.

The normal process of hearing compares most readily with passive sonar. Consider the example of one person listening to another speak: the speaker is the sound source;

the fluid medium is the air; the transducer is the ear, which converts sound waves into a neural signal; and the classifier is the brain of the listener, which extracts useful information from the neural signal. Passive sonar follows a similar pattern: the sound source is a target such as a ship; the fluid medium is the seawater; the transducer is a hydrophone, which converts sound waves into an electrical (digital) signal; and the classifier is a mixture of automated procedures and human decisions.

An active sonar also fits the four-part model set out above; the only difference is that the sound source belongs the sonar system (e.g., a projector or a depth charge) rather than the target. Obvious analogues in the natural world include echolocation used by bats, dolphins and in isolated cases, humans (Au et al., 2000; Stroffregen and Pittenger, 1995). Only passive sonar will receive attention in this thesis; active sonar will not.

Locating Sound Sources

The term *localisation* refers to the process by which a listener judges the distance and angle at which a source is located in relation to the head (Moore, 2004). Two analogous problems in passive sonar are *ranging* (i.e., determining distance) and *target angle estimation*. Passive ranging usually involves triangulation using two well-separated receivers, or timing the difference between the arrival of a sound via a direct path and a reflected path, such as the sea bottom or surface (Waite, 1998). There is evidence that listeners exploit room acoustics in a similar way (in combination with other cues) to judge distance (Mershon and Bowers, 1979).

Modern passive sonars employ an array of receivers to resolve the angle of a target. The bearing of a source in relation to the array can be inferred from the time it takes for a sound pressure wave to travel the distance between hydrophones (Burdic, 1984). Specifically, the constructive and destructive interference caused by the phase lag between receivers at a particular frequency is used to estimate the target angle, in a process called *beamforming*. Human hearing also makes use of two spatially-separated receivers: the ears on either side of the head. *Interaural time differences* assist the location of low-frequency and transient sounds in the horizontal plane (Moore, 2004, Chapter 7). Although the similar strategies employed by sonar and hearing are noteworthy, the issue of source localisation is not considered any further.

Frequency Analysis and Power-based Detection

Frequency analysis, which in this case means breaking a sound up into its constituent frequencies, is a central aspect of both sonar and hearing, and it will be discussed at much greater length in later chapters. It suffices here to mention that the human ear contains a pliable membrane of non-uniform physical constitution, called the *basilar membrane* (BM), which responds to different frequencies at various points along its length. The discrete Fourier transform separates a signal in a similar fashion, namely, by exciting a resonance from frequency bins in response to a sound.

Power-based detection and classification in sonar utilise the DFT power spectrum and disregard the phase spectrum. The total power in the sound (that is, energy per time) is

divided amongst componental frequencies, and particular target sounds are identified by the frequencies they contain. The closest analogue in hearing science is the *place theory*, which states that a sound is principally encoded by the extent to which the basilar membrane is displaced along its length—any other detail besides this is ignored, much as the phase samples of the DFT are ignored in power-based sonar.

Spectral Normalisation

The amount of energy that arrives at the ear (or hydrophone) from a sound source is generally dependent upon its distance from the head (or array), and the energy in the sound may not be evenly-distributed in frequency. Consequently, it is often difficult to decide whether a narrowband source is present at a particular frequency simply by taking an absolute measurement in the spectrum: low energy might indicate a strong signal far away; high energy might indicate a weak signal nearby. The engineering solution encountered most often compares the energy in a signal against its background; that is, it tests whether the energy at a particular frequency exceeds the average energy measured in adjacent bands of the spectrum. This procedure is referred to as *spectral normalisation*.

Sonar processors perform spectral normalisation by dividing the energy of a DFT cell by the average energy of the neighbouring block of cells (or, equivalently, subtraction on a log-scale) (Waite, 1998; Grigorakis, 1997). A comparable effect is brought about in the auditory pathway by *lateral inhibition* amongst nerves cells in the cochlear nucleus (Pickles, 1988). The cells are ordered according to the frequencies at which they respond; they are driven by a log-like compression of the sound energy at that frequency; and the activity of a cell tends to reduce that of its neighbours via inhibitory synapses. This means that contiguous blocks of excited cells tend to mutually suppress each other, whereas tones and spectral edges are enhanced (Shamma, 1985b). This thesis will occasionally refer to spectral normalisation, but it is not a primary concern.

Detection based on Fluctuations in Power

Hearing scientists and sonar engineers have suggested—quite independently, it would seem—that fluctuations in the envelope of a received signal might be used to detect a steady signal against a noise background. The idea is that a clean tonal (i.e., one in which the amplitude remains constant over the period of analysis), when added to a noise signal, will tend to reduce the fluctuation in its envelope. It was along these lines that Wagstaff (1998) developed the *Wagstaff's Integration Silencing Processor* (WISPR) filter. This receiver measures variations in the energy of a DFT time history using the harmonic mean[1]. The DFT bins that capture a steady signal fluctuate less, and this is used to aid signal detection or to distinguish "clutter" signals from stable tonals. The same concept motivated Schoonveldt and Moore (1989) to propose that human listeners use envelope fluctuation (or a lack thereof) as a cue for detecting a tone in noise. This principle extends to across-channel processing in *comodulation masking*

[1] that is, $E\{X^{-1}\}^{-1}$.

release (Hall et al., 1984). In this work, we are principally concerned with the effect of a steady signal upon fluctuations in fine timing structure rather than the envelope.

Detection based on Timing

Some researchers contend that the auditory system encodes information in the time intervals between the discharges of phase-locked fibres, which are then used in pitch analysis (Meddis and Hewitt, 1991) or signal detection (Moore, 2004, page 98):

> "A tone evokes neural firings with a well-defined temporal pattern; the time intervals between successive nerve spikes are close to integer multiples of the period of the tone. A noise evokes, in the same neurons, a much less regular pattern of neural firings. Thus a tone may be detected when the nerve fibres responding to it show a certain degree of temporal regularity in their patterns of firing."

In sonar analysis, Higgins (1980), citing no inspiration from the auditory system as such, attempted to use the steadying effect of a tonal signal upon the zero crossings of band-pass filtered signal as a means of detection. This is the temporal analogue of the envelope fluctuation-based approach described above.

Higgins' experiment required the detection of a 9.9 kHz sinusoidal signal mixed with Gaussian noise, using the zero crossings recorded in the output of a linear filter centred on 9.9 kHz with a bandwidth of 5 kHz. The study concluded that the technique was only effective when the signal-to-noise ratio (which lacked a clear definition in the paper) was somewhat greater than -14 dB.

1.1.4 Auditory-motivated Passive Sonar

Several auditory-motivated sonar algorithms have been developed in recent years. Unlike the examples cited above, these represent attempts to reproduce the success of the biological ear in nature, and the success of imitative technologies (e.g., front-ends for automatic speech recognition), in the domain of passive sonar.

Teolis and Shamma (1991), motivated by anecdotal evidence that human listeners are better able to identify transients than automated techniques based on the power spectrum, developed a sonar transient classifier based on a computational model of the human auditory system. The front-end simulates the auditory periphery using a wavelet transform, sigmoidal compression, spatial filtering and extrema sampling. The specific model used was that of Yang et al. (1991). The noise robustness of the signal representation it generates was subsequently investigated by Wang and Shamma (1992).

A separate study that combined an auditory front-end with a machine learning classifier was carried out by Parks and Weisburn (1992). Two sets of features were extracted: one from the output of a constant-Q filterbank, which served as a cochlear model, and the other from the short-time Fourier transform. Fisher's linear discriminant was then applied to each feature set with the goal of classifying bowhead whale calls and ice

sounds recorded in the Arctic. The results obtained using the cochlear model compared favourably with those obtained using a power-spectrum model.

Tucker and Brown (2005) described an auditory-motivated transient classifier which operates on three perceptually-motivated features: timbre (the quality of the sound), physical material, and temporal context (e.g., any rhythmic pattern surrounding the transient). The underlying acoustic features that listeners use when assessing timbre are discovered using a *multi-dimensional scaling technique* (Grey, 1977). Acoustic features relating to the material of the source are computed from the rate of exponential envelope decay measured for peaks in the output of a cochlear filterbank. The temporal context of the transient is characterised using features extracted from a multi-scale *rhythmogram* representation (Todd, 1994).

Bleeck et al. (2008) have recently presented an application of the *auditory image model* (AIM) (Patterson et al., 1995) in active sonar, in which the size and shape of an object is determined from its simulated scattered signal. The AIM extracts a stabilised timing representation from the response of a population of model fibres using autocorrelation (reviewed in Section 2.2.4). A similar active sonar classification study, which uses an auditory model to analyse musical timbre, has been conducted by Young and Hines (2007).

Finally, Pykett and Smith (2000), in a report examining how the ear, as a "highly successful biological signal processing system", might aid the design of passive sonar, suggest using the fine timing structure of harmonics as a cue for grouping them:

> "In passive sonar it is important to be able to assign frequency lines in a time-frequency display to membership of various harmonic series. These can then be related to acoustic generating mechanisms in a distant target. Because of the low signal-to-noise ratios frequently encountered[,] it is sometimes difficult to identify all the components which belong to a particular series. [...]
>
> [A model of the auditory system] produces spikes with intervals which are multiples of the period of a harmonic component of the signal. For coincidence to occur with the output of the oscillator circuit, modulation (fundamental) and carrier periods much match. Which harmonic relates to a particular fundamental, can then be determined by the coincidence neuron which fires. [...] The ear seems to have struck an effective balance for its purpose, military systems may require different tuning."

The sixth objective of the thesis (*cf.* §1.3.1) pursues this idea.

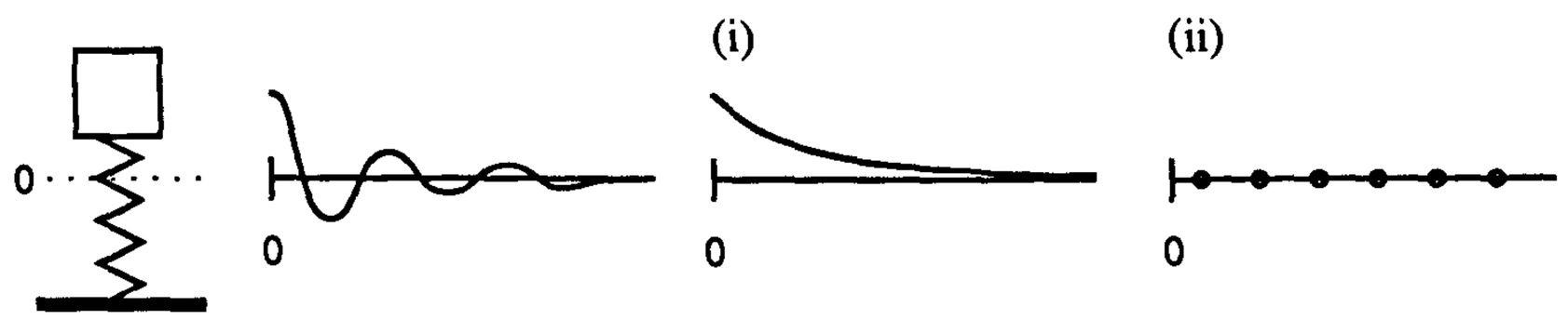
1.2 Assessing the Temporal Structure of a Signal

1.2.1 Temporal Coding

A temporal code may be defined for our purposes as "a principled method for encoding information by non-uniformities in time". Many codes—written, biological or digital—encode information by *varying the symbols* in *fixed time slots*. A temporal code, by contrast, encodes information by the *varied timing* of *fixed symbols*. Acoustic signals can be characterised either way, and it is a matter of debate which kind of code the auditory system employs.

An Illustration

Consider a mass loaded on vertical spring, which is under the influence of gravity, like a Jack-in-the-box. If the mass is held at a certain height and released, then, assuming no losses, the energy in this system will remain constant, being perpetually converted between potential energy in the spring and kinetic energy in the mass' motion. If the mass encounters resistance proportional to its speed, the energy in the system will gradually drain away (Kinsler et al., 2000).



A graph of energy against time, as in (i) above, does not convey all the pertinent information about the system. For instance, it omits the frequency at which the mass bobs up and down. If, instead, we imagine that a switch is closed every time the mass passes its point of equilibrium (chosen to be zero here), and plot the output of the switch, then we obtain the kind of point process shown in (ii), which captures the upward and downward motion of the mass. Notice that it is the distribution of one symbol (o) in time that conveys information.

Temporal Theories of Auditory Encoding

The vibration of a section of the basilar membrane in response to a stimulus may be likened to the mass-loaded spring scenario above. A row of sensitive hair cells line the BM and these, like the switch on the spring, are depolarised at a particular phase of the membrane motion, causing a "spike" to cascade along the auditory nerve to the brain. If every cell discharged consistently, then the information communicated to the brain would resemble mode (ii) above: the temporal information would be entirely preserved, the envelope entirely lost. However, this tidy picture is complicated by the stochastic nature of cell firing.

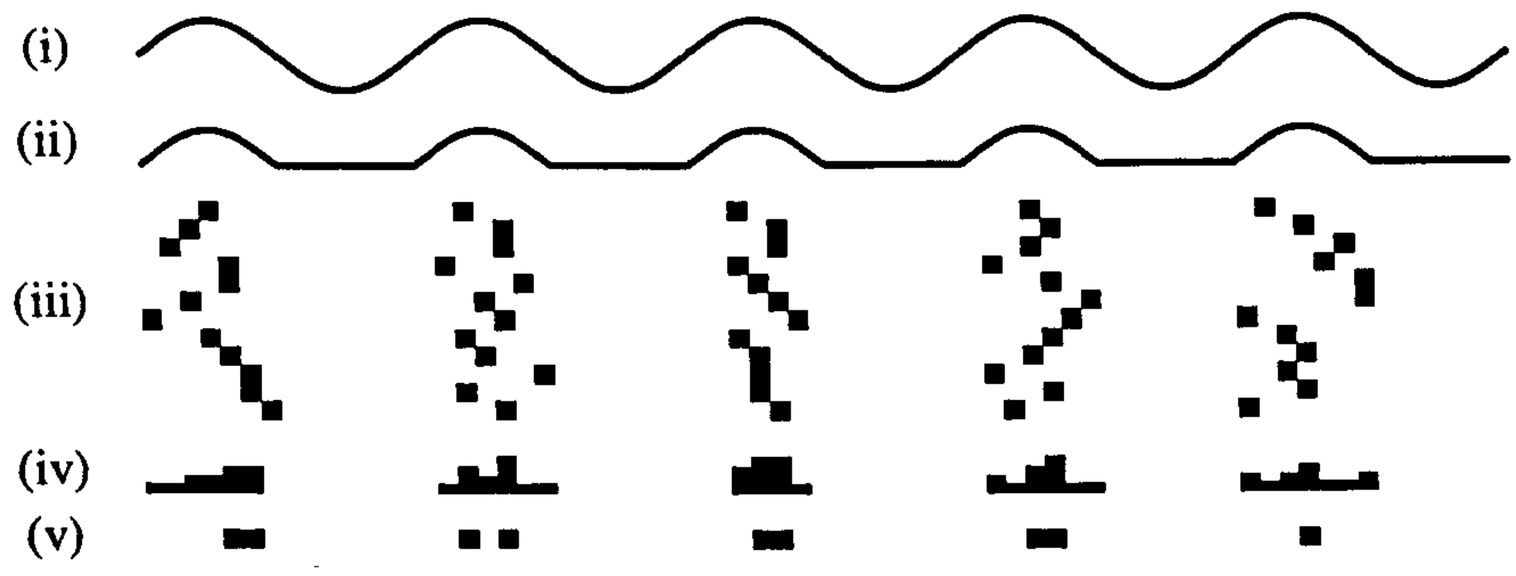


Figure 1.1: The volley principle applied to randomly-generated data: (i) BM motion; (ii) half-wave rectified BM motion; iii) twelve noisy spike trains, where the spikes are distributed over the half-period in proportion to the BM amplitude; iv) net response; v) spikes generated when at least 1/4 of the population fire. The spike train in (v) is almost phase-locked to the peaks. Integrating more 'fibres' would further improve the response.

First, some fibres are more likely to discharge when the displacement of the BM is greater, leading some researchers to suggest that the net response of a *population* of hair cells encodes displacement (a correlate of the envelope). This rate coding theory, in its strongest form, denies the role of timing altogether. Second, fibres phase-lock to the stimulus in an approximate fashion, which means that individual spike trains communicate a rather degraded copy of the BM motion. The *volley theory*, proposed by Wever (1949), allows the fine structure to be recovered from many noisy spike trains by a process of integration, as Figure 1.1 illustrates. The relative merits of various coding theories are discussed in Chapter 2.

1.2.2 Mathematical Analysis of Temporal Codes based on Zero Crossings

Consider the task of detecting a weak sonar signal embedded in noise. If we are to design timing-based algorithms capable of competing with power-based algorithms in this regard, we must choose a suitable set of mathematical tools for interpreting the zero crossings of a signal. Much as the Fourier transform supplies a mathematical "bridge" for moving back and forth between the time and frequency domains, this section describes three frameworks for relating the samples of a signal to its zero crossings, for switching between non-temporal and temporal codes.

Product Representations

A *product representation* describes a signal entirely in terms of one scale factor and a set of zeroes in the complex time domain. This kind of representation has been successfully applied to spectral analysis (Kay and Sudhaker, 1986) and positive instantaneous frequency estimation. In the latter case, the literature mentions auditory inspiration (Kumaresan and Wang, 2001) and possible applications in sonar (Kirsteins et al., 2000), making this mathematical framework particularly worth exploring.

The transformation of a time domain signal into a product of *elementary signals* can be described in three steps. First, a band-limited signal with period T, $g(t)$, is expanded into a Fourier series with complex coefficients $\hat{X}[s]$, as follows

$$g(t) = \sum_{s=-S}^{S} \hat{X}[s] \exp\left(\frac{i2\pi st}{T}\right). \tag{1.1}$$

The highest frequency (or one-sided bandwidth) of the signal is dictated by S. Second, using the abbreviation $\xi = \exp(i2\pi t/T)$, it is possible to write (1.1) as a polynomial in ξ of degree $2S$, i.e.,

$$g(t) = \xi^{-S}\left[\hat{X}[-S] + \hat{X}[1-S]\xi + \hat{X}[2-S]\xi^2 + \cdots + \hat{X}[S]\xi^{2S}\right]. \tag{1.2}$$

Note that the time parameter t can be complex, and evaluating $g(t)$ along the real axis recovers the original signal. The fundamental theorem of algebra guarantees that a polynomial of degree $2S$ can be factorised into a product of $2S$ complex roots, $r_1, \ldots, r_{2S}$. This leads to the third step: replacing the square-bracketed portion of (1.2) with a product:

$$g(t) = r_0\,\xi^{-S}(\xi - r_1)(\xi - r_2)\cdots(\xi - r_{2S}) \tag{1.3}$$

$$= r_0\,\xi^{-S}\prod_{s=1}^{2S}(\xi - r_s), \tag{1.4}$$

where $r_0 \equiv \hat{X}[S]$. (1.4) is known as a product representation (Voelcker, 1966). Note that zeroes occur in the real (i.e., visible) signal for all $\xi = r_s, |r_s| = 1$.

This representation provides some fascinating insights into the nature of a signal, which are not apparent in the more traditional domains. These include the conditions under which a signal can be reconstructed from its zero crossings (Logan, 1977), techniques for modifying a signal so that its spectrum can be decoded from its zero crossings (Kay and Sudhaker, 1986), and methods for decomposing a signal into positive instantaneous frequency (PIF) and envelope signals (Kumaresan and Rao, 1999). A product representation can decompose a deterministic signal in useful ways, but it is not clear how readily it can be applied to random signals.

Perturbation Analysis

The *perturbation* interpretation of zero crossings considers the effect of weak, Gaussian noise on the zero crossings of a deterministic signal $s(t)$ and has been adopted by a number of workers (Park and Stern, 2006; Sekhar and Sreenivas, 2005; Kim et al., 1999; Sreenivas and Niederjohn, 1992). The noise samples are drawn from a Gaussian process n with a slow-varying envelope, zero mean and variance σ_{n}^2. The received signal is modelled as the sum of the signal and noise:

$$g(t) = s(t) + \mathsf{n}. \tag{1.5}$$

Suppose that the clean signal contains a zero crossing at time t_s. If we assume that the signal $s(t)$ is smooth through its zero crossings, we can approximate the waveform

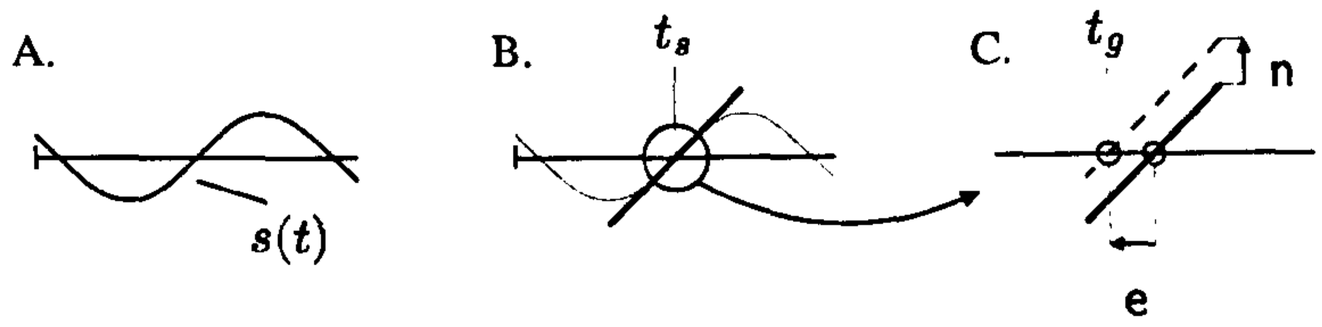


Figure 1.2: A) a short section of the clean waveform, $s(t)$; B) the Taylor series expansion around a zero crossing of $s(t)$ at time t_s; C) the addition of a constant (n) to this line shifts the axis crossing by a proportional amount (e).

around t_s using a first-order Taylor series expansion.

$$\tilde{s}(t) = s(t_s) + \left.\frac{ds}{dt}\right|_{t=t_s} (t - t_s) \tag{1.6}$$

$$= \left.\frac{ds}{dt}\right|_{t=t_s} (t - t_s). \tag{1.7}$$

Now (1.7) is the equation for a line to which we shall add a constant noise sample n. Label the perturbed zero crossing time in the noise-added signal using t_g. Then,

$$g(t_g) \approx \left.\frac{ds}{dt}\right|_{t=t_s} (t_g - t_s) + \mathsf{n} = 0. \tag{1.8}$$

Rearranging (1.8), we observe that the perturbation (i.e., small difference) in the zero crossing time, e, is directly proportional to the additive noise sample:

$$\mathsf{e} \triangleq (t_g - t_s) = -\left[\left.\frac{ds}{dt}\right|_{t=t_s}\right]^{-1} \mathsf{n}. \tag{1.9}$$

As n is Gaussian, e is also Gaussian, with zero mean and variance

$$\sigma_{\mathsf{e}}^2 \left[\left.\frac{ds}{dt}\right|_{t=t_s}\right]^2 = \sigma_{\mathsf{n}}^2.$$

Having shown that the perturbation of a zero crossing is Gaussian, it is a short step to demonstrate that the perturbation of many zero crossings in the presence of weak white or coloured Gaussian noise is governed by a joint Gaussian distribution, as are the intervals between the crossings.

This kind of mathematical treatment, though successful in other domains, is ill-suited to our purpose: firstly, because it refers primarily to the effect of noise upon a known signal, which we do not possess; and secondly, because the assumption that axis crossings are linearly-perturbed about a steady state requires a high SNR, which can by no means be guaranteed at sea.

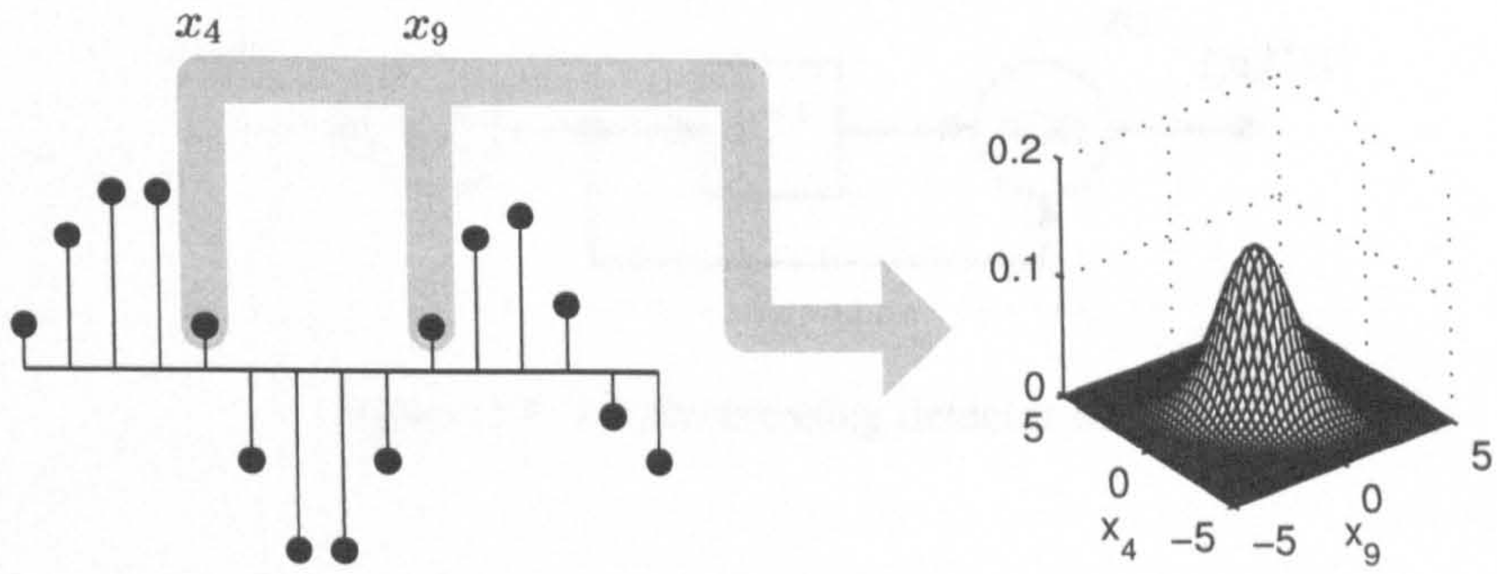


Figure 1.3: The stem plot on the left shows one possible sample function of a Gaussian process. Samples x_4 and x_9 are governed by a two-dimensional Gaussian distribution.

Sign Changes in a Random Process

Finally, zero crossings may be interpreted as sign changes in a random process from negative to positive or *vice versa*. The ideas reviewed next provide the framework in which the majority of the work in the thesis will be carried out.

A *Gaussian process* is any random process for which all possible subsets of samples are governed by a joint Gaussian distribution, and a number of zero crossing-related results are available in this special case. For instance, Figure 1.3 illustrates how a pair of samples values, x_4 and x_9, are modelled as a bivariate Gaussian. In general, n samples are modelled by an n-variate Gaussian distribution.

A *wide-sense stationary* random process is any process for which the mean sample value is a constant, and the covariance of pairs of samples is time-invariant, depending only on their separation. This implies that if the process depicted in Figure 1.3 were wide-sense stationary, then the density shown would govern not only (x_4, x_9), but also (x_0, x_5), (x_1, x_6) and, in general, (x_n, x_{n-5}). A zero mean, wide-sense stationary Gaussian process, X, is characterised entirely by its *autocovariance function*,

$$\gamma_X[k] = E\{x_n x_{n-k}\}. \tag{1.10}$$

For continuous-time processes we write $\gamma(\tau)$, where τ is a time lag. In general, square brackets [ ] and round brackets ( ) are used to indicate functions of a discrete argument and continuous argument, respectively.

Kedem (1980) provides a derivation of the probability of a zero crossing in terms of the changes of sign in a wide-sense stationary Gaussian process. Pairs of consecutive samples, which for simplicity's sake we shall label x_1 and x_2, are governed by a bivariate Gaussian distribution. The probability of observing a zero crossing on a sample in either direction is therefore given by

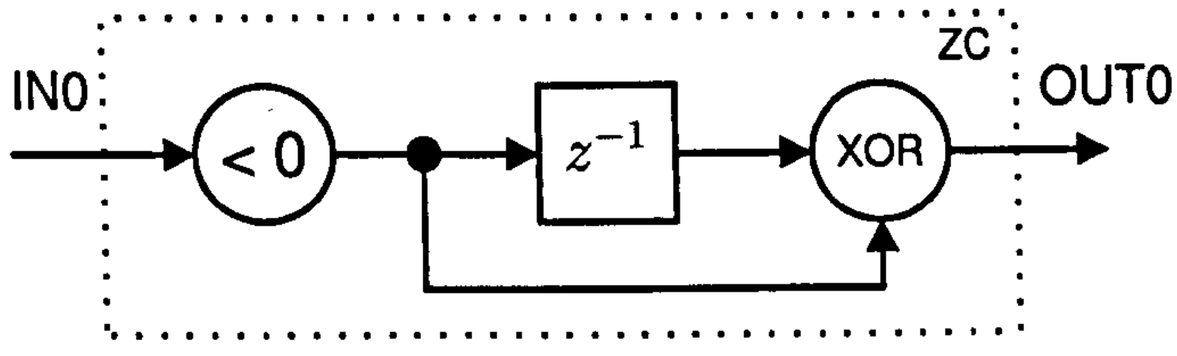


Figure 1.4: A zero crossing detector block.

$$
\begin{aligned}
& P(C) \\
&= P(x_1 \geq 0, x_2 < 0) + P(x_1 < 0, x_2 \geq 0) && (1.11) \\
&= 2 \times \frac{\gamma^2[0] - \gamma^2[1]}{2\pi} \int_{-\infty}^{0} \int_{0}^{\infty} \exp\left[\frac{\gamma[0]x_1^2 + \gamma[0]x_2^2 - 2\gamma[1]x_1x_2}{-2(\gamma^2[0] - \gamma^2[1])}\right] dx_1 dx_2 \\
&= \frac{1}{2} - \frac{1}{\pi} \sin^{-1} \rho[1], && (1.12)
\end{aligned}
$$

where $\rho[k]$ denotes the *autocorrelation function*, which is the autocovariance function normalised by the variance[1], i.e., $\rho[k] \triangleq \gamma[k]/\gamma[0]$.

Consider the zero crossing detector block diagram in Figure 1.4. If the input at `IN0` is a wide-sense stationary, zero mean Gaussian process, and $\rho[1]$ is known, then the probability that `OUT0` outputs 1 on a given time step can be found using (1.12). Furthermore, if the input signal is sampled at a rate of f_s samples per second, then the expected number of output "spikes" per second is

$$\frac{f_s}{2} - \frac{f_s}{\pi} \sin^{-1} \rho[1].$$

If we imagine that the input signal is sampled at ever higher rates, then we can obtain a similar result for a continuous signal by taking a limit[2]:

$$\lim_{f_s \to \infty} \left\{ \frac{f_s}{2} - \frac{f_s}{\pi} \sin^{-1} \rho\left(\frac{1}{f_s}\right) \right\} = \frac{1}{\pi} \sqrt{\rho''(0)}. \qquad (1.13)$$

The right-hand side of (1.13) relates the expected number of zero crossings in unit time for a zero mean, wide-sense stationary Gaussian process, and it is frequently referred to as *Rice's Formula*, after S. O. Rice, who included it in his monumental work, *Mathematical Analysis of Random Noise* (Rice, 1944).

[1] Rather unhelpfully, the terms "autocovariance" and "autocorrelation" are often used interchangably in the literature. It must be emphasised that auto*covariance* here refers to the covariance of pairs of samples; auto*correlation* refers to their Pearson product-moment correlation coefficient (i.e., $-1 \leq \rho \leq 1$).

[2] Here, $\rho'(\cdot)$ and $\rho''(\cdot)$ are used to denote the first and second derivative of the autocorrelation function, respectively.

1.3 Thesis Overview

1.3.1 Objectives

Temporal Analysis of Sonar Signals

The acoustic signal arriving at a single, omnidirectional hydrophone is an additive mixture of target signal and background noise. Conventionally, the ratio of signal power to noise power is the quantity of chief importance in sonar system performance analysis, and most components of the system can be viewed as attempts to improve it. Beamforming reduces the noise power arriving from unwanted angles; Fourier analysis reduces the noise power contributed in unwanted frequency bands; care is even taken to ensure that the window function applied prior to the DFT improves SNR.

The human ear is not exclusively a power detector. According to temporal theories of encoding, patterns in the discharge times of auditory nerve fibres communicate information to the brain; that the phase-locking of fibres has at least the *potential* to encode features of a stimulus is uncontroversial.

(1) Can timing-based auditory models be adapted to perform narrowband sonar analysis? What benefits might this offer?

Statistical Timing-based Detectors

One of the advantages of power-based detection is the existence of relatively simple statistical models for commonly-encountered random processes such as Gaussian noise and sinusoids (Whalen, 1971). This enables the construction of theoretical receivers, whose performance under certain conditions can be determined analytically, e.g., the rate of false alarm, or the sensitivity required to secure a 10% chance of detection. A pixel on a narrowband display can be tested for a signal in a principled way, if the transformations that the signal undergoes between the hydrophone and the display (or even its viewer) are correctly characterised.

The second aim is to develop a suite of *ideal temporal receivers*: algorithms that operate on the fine structure of a signal rather than its envelope, and for which an input-output relationship can be formulated in statistical terms. For example, suppose there is a 50% chance that a unit-amplitude sinusoid has been added to a Gaussian noise background with unit variance. If a series of zero crossing measurements are taken, what is the theoretically-optimum decision concerning whether the sine wave is present? If we choose accordingly, what is the probability we are wrong? How many more zero crossings must we measure before there is a 95% chance we are right? Ideal temporal receivers will allow us to answer such questions.

(2) How does the performance of an elementary interval detector, which operates on one zero crossing interval, compare to that of a power detector, which operates on one sample of the envelope?

Hopefully, the elementary interval detector will elucidate the mechanisms at work in the production of a timing-based display, much as the quadrature receiver provides theoretical support for more elaborate power-based displays. If this detection model is constructed successfully, then the following two questions must be addressed.

(3) Is it possible to develop a hybrid detector, which uses both power and timing information? Do a sample of the squared-envelope and a zero crossing interval convey mutually-exclusive or equivalent information?

(4) Can the elementary interval detector be modified to incorporate multiple interval observations, analogous to a spike train?

Adapting Computational Auditory Scene Analysis Methods to Sonar

The first four objectives concern the design of sonar algorithms based on timing models of the peripheral auditory pathway. The remaining objectives take their inspiration from auditory scene analysis (ASA)—higher-level organisational principles, which govern how a listener groups components of a signal into perceptual wholes (Bregman, 1990). The latter part of the thesis is driven by the broad aim of discovering a set of organisational principles applicable to sonar and designing algorithms to implement them. Two aspects of auditory scene analysis will receive particular attention, however.

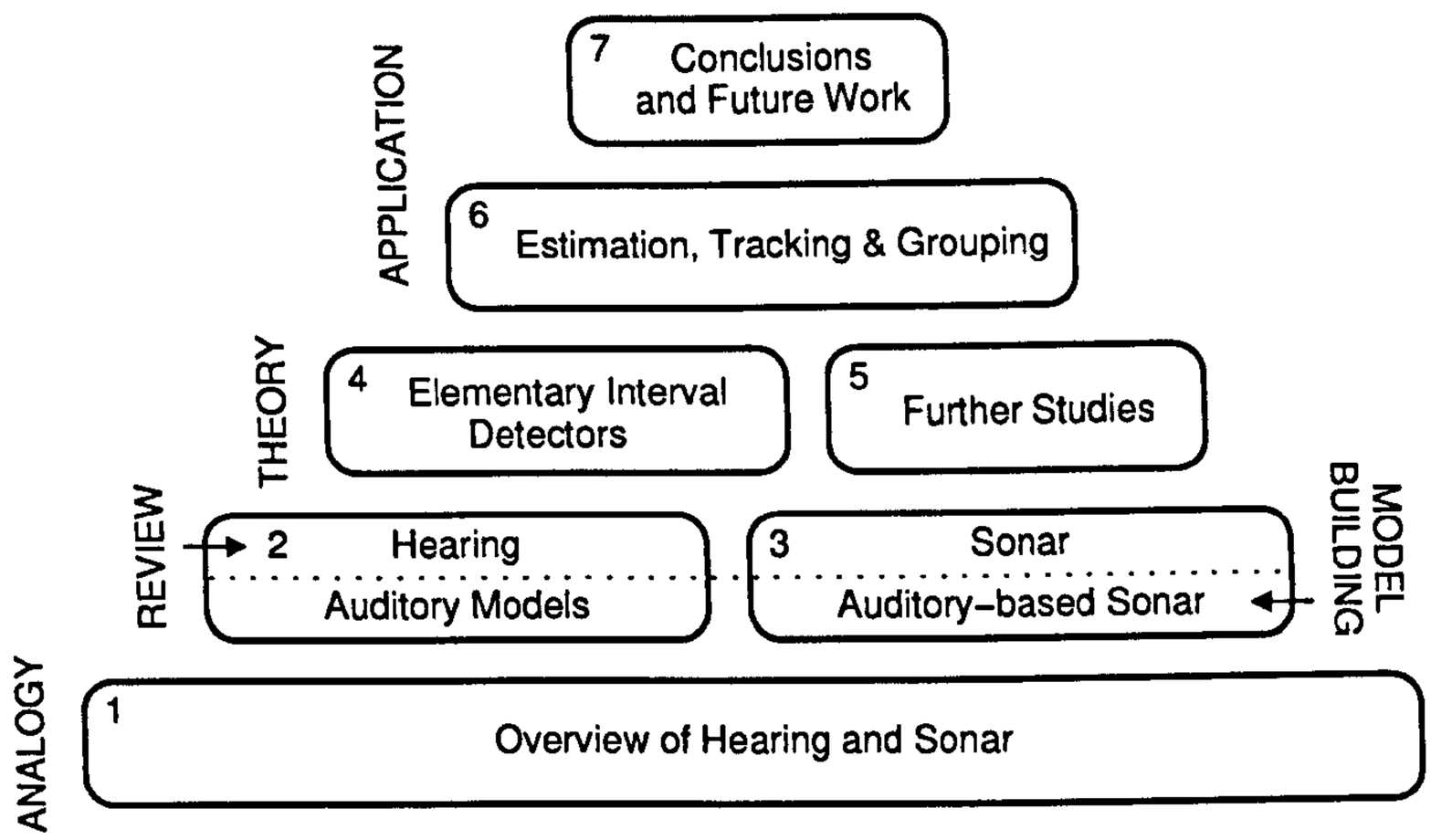
(5) ASA causes a listener to perceive the continuation of a tone which is masked momentarily by noise. Can a similar principle be used in sonar to reconstruct a tonal interrupted by a transient event?

(6) ASA promotes the fusion of partials exhibiting a common pattern of modulation, especially those in a harmonic relationship. Can a similar principle be used to group engine tonals?

1.3.2 Scope

Having outlined the objectives of the thesis, we must state a couple of caveats. Although these algorithms draw inspiration from the physiology of the ear, there is no intention of closely modelling the ear in every respect. The general strategy will be to start out with an auditory-scale model and then adjust it so that it can be applied in realistic conditions. The human ear lacks the frequency-resolving power to carry out the kind of narrowband analysis required of a sonar receiver. Accordingly, in the detection tasks used to evaluate the timing-based detectors, the goal is not to match the performance of a human listener, but rather that of a power detector.

This leads to the second caveat, which concerns optimality. When referring to the "optimal detector" for a particular measurement (e.g., the squared-envelope or a zero crossing interval), we mean that, from the set of all detectors which utilise this quantity, this detector maximises some performance criterion. It should not be taken to imply that it is impossible to construct a better detector *of any kind*. Whalen (1971, page 155) helpfully defines the optimal receiver as "a receiver which best satisfies a given criteria [*sic*] under a given set of assumptions."



1.3.3 Structure

The remaining content of this thesis is divided across five chapters of approximately equal length, and the major findings are summarised in a final chapter. The entire thesis has been organised into the pyramidal structure drawn above. The internal structure of each chapter is described in its opening paragraphs.

Conscious that the readership of this thesis will be drawn from two possibly disjoint camps—hearing scientists and sonar engineers—a substantial portion has been devoted to **review**: the second layer of the pyramid. Chapter 2 provides an overview of auditory physiology and psychology in its first half and describes computational models of hearing in its second half—particularly those based on timing. Chapter 3 provides an introduction to sonar and considers which of the auditory models in Chapter 2 might be most-readily adapted for sonar analysis. By the end of the third chapter, Objective 1 should have been addressed.

Chapters 4 and 5, in layer three, concern the **theory** of ideal temporal receivers and attempt to address Objectives 2 and 3–4, respectively. Chapter 4 is restricted to very simple detectors (single intervals only) and noise models (stationary Gaussian processes). Chapter 5 moves beyond these restrictions to consider more advanced detectors (multiple intervals, joint detection based on timing and power) and noise models (sinusoids with constant and Rayleigh amplitude).

The primary interest of Chapter 6 (and to some extent, Chapter 7) is **application**. One goal is to unify the material in the layers underneath. For instance, Chapters 4 and 5 were written in response to the need for a formal understanding of the timing-based sonar displays in Chapter 3. This chapter will apply these theoretical findings to the practical design of sonar displays where possible. A second goal is to examine whether the auditory scene analysis techniques described in Chapter 2 can be incorporated into the new, timing-based sonar algorithms, in order to fulfil Objectives 5 and 6.

Chapter

2

Physiology, Psychology and Computer Models of the Ear

The ear is the most sophisticated audio signal processor known to man. It is also the chief source of inspiration for novel sonar signal processing techniques undertaken in this study and as such merits detailed examination. This chapter is structured as two sections in sequence. The first section focuses on the physiology of the ear and, to some extent, the psychology of the listener; the second section, building on the first, reviews the computational models that have been proposed to explain or reproduce various aspects of human hearing for scientific, clinical or technological ends. There are also parallel ties bridging the subsections: the first half is composed of subsections that trace an orderly progression from the exterior of the head (periphery) to the brain (centre); the discussion in the second half is organised so as to mirror the first half.

Chapter 2 Outline

	Physiology / Psychology (2.1)	Computational Modelling (2.2)
peripheral	Outer–Middle Ear (2.1.1)	Outer–Middle Ear Models (2.2.1)
	Cochlea (2.1.2)	Basilar Membrane Models (2.2.2)
	Auditory Nerve (2.1.3)	Transduction Models (2.2.3)
	Theories of Encoding (2.1.4)	Models of Encoding (2.2.4)
central	Auditory Scene Analysis (2.1.5)	Computational ASA (2.2.5)
	Interim Summary (2.1.6)	
	physiology	model

This chapter accomplishes two stated aims. All auditory modelling studies necessarily make use of abstraction—the notion that auditory function is explicable at the level of the effective signal processing it performs. By the end of this chapter, I intend to have collected a number of "effective signal processing methods", each with some identifiable physiological basis, that can be carried over into Chapter 3 and evaluated with respect to their suitability for sonar applications. The second aim of the chapter is to isolate and critically investigate those aspects of auditory physiology and modelling that emphasise the use of fine temporal structure in signals, generally expressed—at least, in modelling terms—in the phase or autocorrelation of narrowband signals. Timing is an additional theme that will persist into the third chapter and beyond.

2.1 Physiology and Psychology

The sense of *hearing* enables an organism to explore its environment by analysing pressure waves. Identifying surrounding objects by the sounds they emit is one major aspect of this analysis and the focus of this section. The material is presented in five parts and expatiates the following brief description of the hearing process. First of all, the (typically) airborne sound is transmitted to fluid inside the cochlea (§2.1.1), then pressure in the fluid displaces flexible structures within the cochlea in such a way that certain parts move in response to particular frequencies and activate adjacent nerve fibres (§2.1.2). The response of an individual nerve fibre captures numerous properties of a simple sound stimulus (§2.1.3), and there are various theories regarding how a population of nerve fibres may encode a complex stimulus like speech (§2.1.4). Finally, auditory scene analysis is a theoretical framework that attempts to identify the principles by which a neural signal is organised if many sound sources are heard together (§2.1.5). For a comprehensive treatment of the physiology and psychology of hearing, the reader is encouraged to consult Pickles (1988) and Moore (2004), respectively.

2.1.1 The Outer and Middle Ear

Auditory processing in humans commences at the outer ear, which consists of the auricle (*pinna*) and external auditory canal (*meatus*). The auricle is the visible part of the ear that projects from the side of the head, and its principal role is to direct sound waves arriving at the head into the auditory canal, an air-filled duct leading to the eardrum or *tympanic membrane* (Pickles, 1988). In this way, sounds in the environment are reproduced as vibrations in the eardrum. The outer ear also serves to modify sound pressure at the tympanic membrane and assist in the spatial localisation of sound sources.

The structures of the middle ear are located in a cavity between the eardrum and the *oval window*. The vibration of the eardrum is communicated to the oval window via three small, interlocking bones, referred to individually as *incus*, *malleus* and *stapes*, and collectively as *ossicles*. The purpose of the ossicles is to overcome the difference in acoustic impedance between the air in front of the eardrum and the fluid behind the oval window, the entrance to the cochlea. Were no such mechanism present, most of the energy in a wave incident upon the air-fluid boundary would be reflected.

2.1.2 The Cochlea

The cochlea is a distinctive snail shell-shaped cavity, which forms part of a larger network of fluid-filled canals called the inner ear. The cochlea is an integral component of the human auditory periphery, and its functions include both the frequency analysis of the vibrations received from the the middle ear and the transduction of those vibrations into a neural signal. It is customary to refer to the cochlea as though it were rolled out flat, labelling the ends corresponding to the edge and centre of the spiral the 'base' and 'apex', respectively.

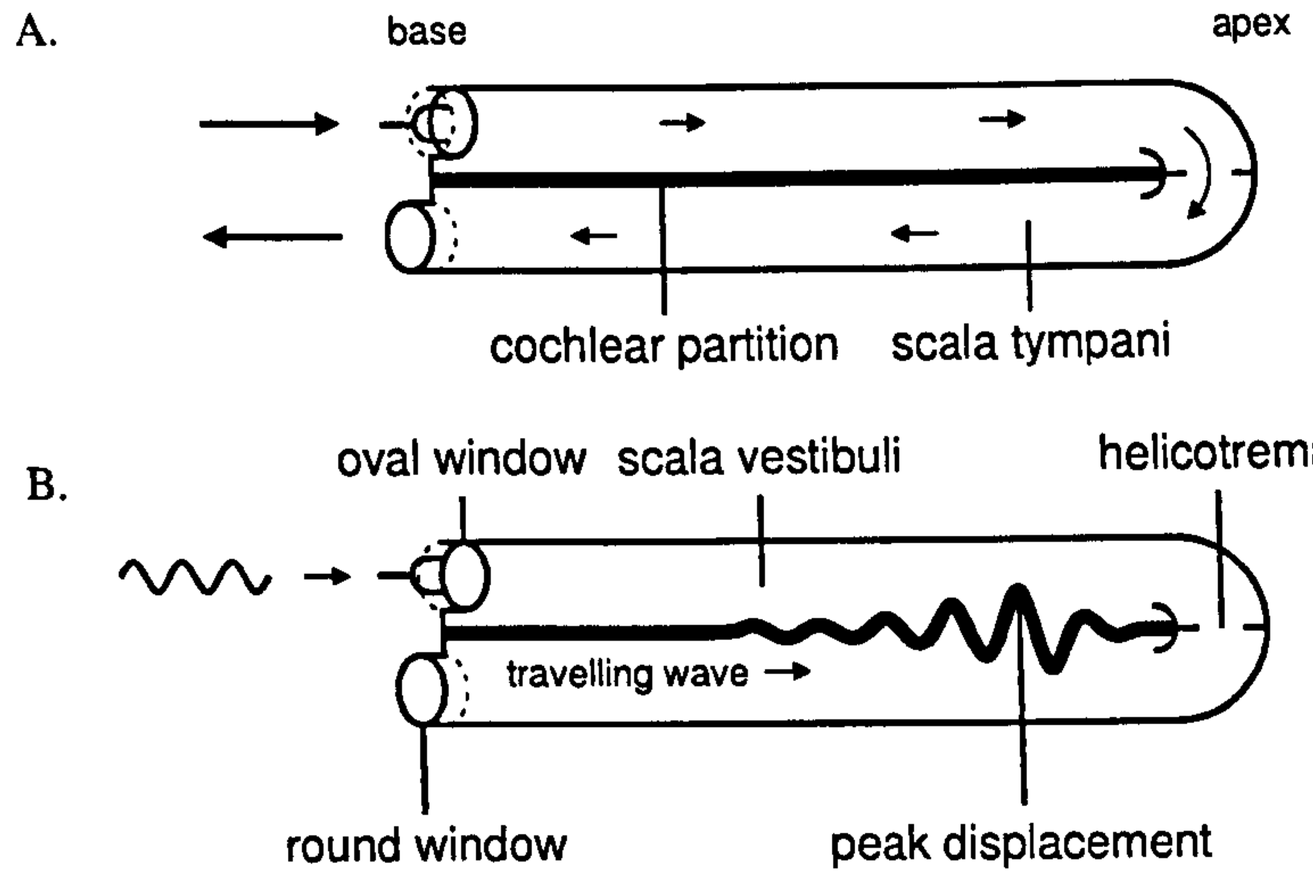


Figure 2.1: A diagram of the cochlea. A) pressure applied to the oval window is transmitted through the cochlear fluid and released at the round window. B) the pressure difference between the scalae sets up a travelling wave in the the cochlear partition which, for a sinusoidal stimulus, reaches a peak displacement near its place of resonance.

The cochlea is partitioned longitudinally into three chambers by two membranes: Reissner's membrane and the basilar membrane. The two outer chambers are called the *scala vestibuli* and *scala tympani*, and the chamber situated between the two membranes is called the *scala media*. A small aperture at the apical end of the cochlea, the *helicotrema*, permits the passage of fluid between the two outer chambers. At the base of the cochlea are located two membranes: the oval window, which projects onto the scala vestibuli, and to which the stapes of the middle ear adheres; and the *round window*, which projects onto the scala tympani.

The vibration of the stapes is transmitted through the oval window and produces a longitudinal wave in the cochlear fluid, which propagates along the scala vestibuli and scala tympani and finds a pressure release at the *round window*, as Figure 2.1A shows. As this wave progresses, it creates a pressure difference across the *cochlear partition*—Reissner's membrane, the basilar membrane and intervening structures—causing it to move. If the stimulus applied to the oval window is a sinusoid, the entire cochlear partition vibrates at the stimulus frequency. However, the phase and amplitude of a vibrating point on the cochlear partition varies as a function of its distance measured from the stapes.

Experiments conducted by von Békésy (1947) demonstrated that the transverse motion of points along the cochlear partition describe a travelling wave, which originates at the base and proceeds towards the apex, as illustrated in Figure 2.1B. As the travelling wave propagates along the basilar membrane, its envelope grows until reaching a peak, or resonance, after which it rapidly diminishes. For a sinusoidal stimulus, the place where the resonance occurs is related to the stimulus frequency. This relationship arises from the mechanical properties of the partition itself: a high-frequency sinusoid

produces a peak displacement near the base of the cochlea, where the membrane is stiff and narrow; conversely, a low-frequency sinusoid produces a peak displacement towards the apex, where the membrane is pliable and broad.

Von Békésy's initial studies of the vibration excited along the basilar membrane by a tone stimulus reported a broadly-tuned response, too insensitive to account for the ear's frequency selectivity. It is now appreciated that these experimental findings were affected by the poor condition of cochleae extracted from human cadavers and the measurement technology of the era (Moore, 2004). Subsequent research has shown that the live, mammalian cochlea incorporates a feedback mechanism, which sharpens the response of the basilar membrane to a sound stimulus. This mechanism is effected, at least in part, by an array of *outer hair cells* (OHC) inside the cochlear partition, each of which reacts to the velocity of the basilar membrane by actively reshaping its body and so influences the displacement magnitude and phase both along and across the basilar membrane (Nilsen and Russell, 2000).

From a simplified perspective, the basilar membrane may be likened to a frequency analyser, resolving the spectral content of a sound along a spatial axis. The resolution of this frequency analysis is not perfect; a mixture of two tones closely spaced in frequency produces in the basilar membrane a single, broad resonance rather than two distinct peaks. However, a listener's ability to discern a difference in frequency between a pair of tones presented *separately* is rather more impressive, with changes as small as 1–2 Hz in a 1 kHz tone being detectable (Wier et al., 1977). The frequency-resolving power of the basilar membrane is also non-uniform; frequency discrimination is most effective in the 500–1000 Hz range and considerably worsens above 4 kHz (Greenberg and Ainsworth, 2006). A comprehensive review of cochlear mechanics is provided by Robles and Ruggero (2001).

Hair Cell Transduction

Like all sensory organs, the ear is responsible not only for reacting to external stimuli but also for converting those stimuli into neural activity—a process referred to as *transduction*. In the auditory system, the movement of the basilar membrane is transduced into nerve activity by *inner hair cells* (IHC), which are distributed along the interior of the cochlear partition. The displacement of the basilar membrane causes a shearing action between the basilar and *tectorial* membranes, deflecting the hairs or *stereocilia* that line the space inbetween. These hairs form part of the *organ of Corti*, a cross-section of which is illustrated in Figure 2.2.

Each stereocilium is attached to either an inner hair cell or an outer hair cell. It is the inner hair cells which accomplish the transduction process; outer hair cells are implicated in the active basilar membrane tuning mechanism mentioned above and do not concern our present discussion. The deflection of IHC stereocilia opens *transduction channels*, releasing positively-charged potassium ions into the cell body, depolarising the cell. Between each deflection is a brief period of recovery, as the transduction channels close, restoring the potential difference across the cell to some degree. Oscillation of the basilar membrane modulates the cell voltage and steadily depolarises the cell. Each depolarisation of the cell is accompanied by the release

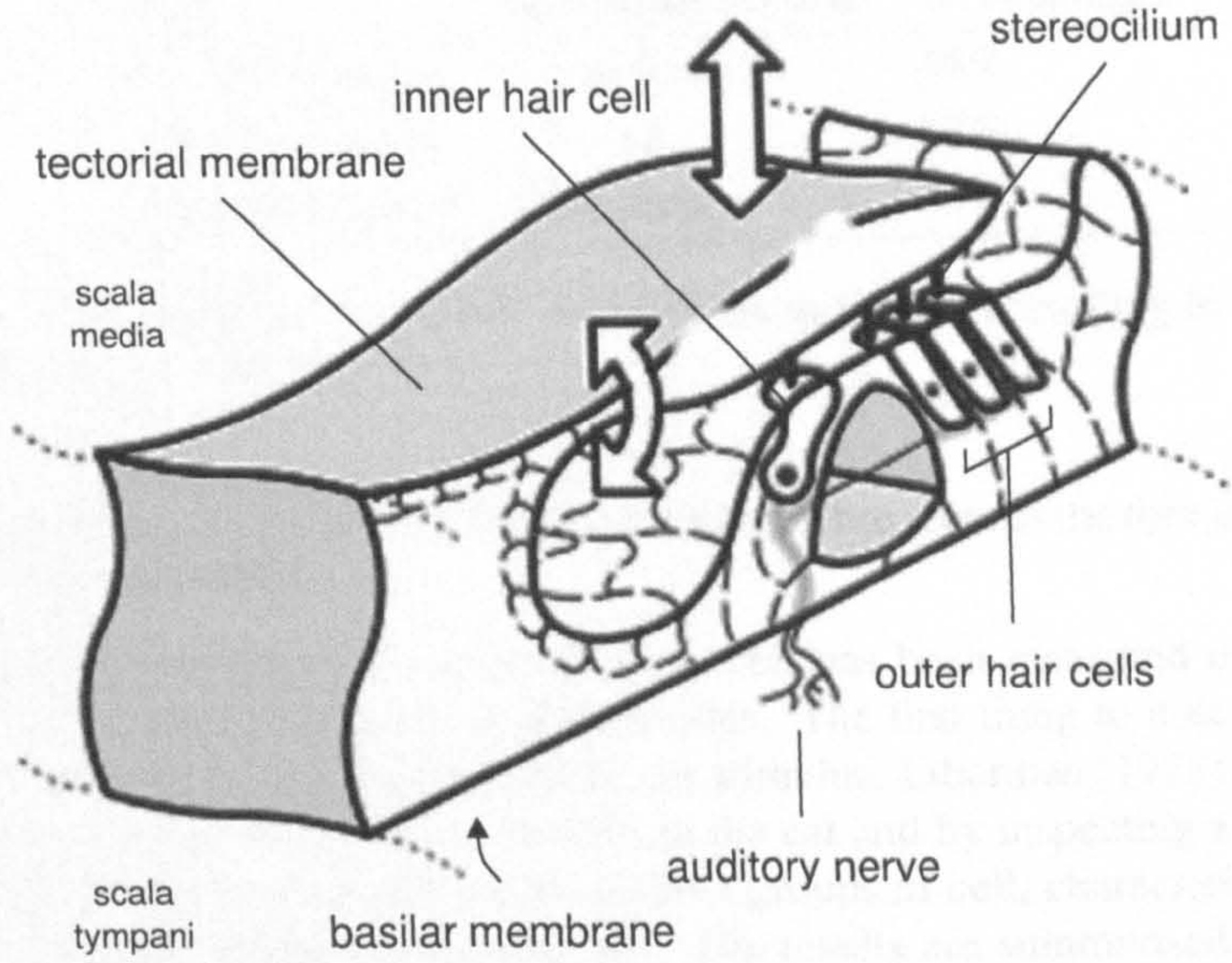


Figure 2.2: A diagram showing a cross-section of the organ of Corti. The tectorial membrane is free to move up and down in relation to the basilar membrane, to which it is 'hinged' on one side (here, the left). When the two membranes are pressed together, the inner hair cell is activated and is more likely to transmit a spike along the auditory nerve.

of neurotransmitter into the cleft between the IHC and the innervating auditory nerve fibre. A sufficient quantity of neurotransmitter evokes an action potential or *spike* in the latter, which is communicated to the brain via the auditory nerve.

2.1.3 The Auditory Nerve

We have examined thus far the various stages in which an acoustic wave arriving at the outer ear is transformed into spiking patterns in the auditory nerve cells innervating the cochlea. Physiological research in the last half-century has greatly contributed to our understanding of how the properties of a tone stimulus—its frequency, intensity and phase—are represented in the discharge patterns of individual fibres. Having reviewed the key results of these studies, we will be ready to discuss how a population of fibres might encode a complex stimulus, such as speech or music.

Average Firing Rate and Spontaneous Activity

The response of individual auditory nerve fibres to a stimulus was first investigated by Tasaki (1954) and remains the subject of ongoing research today. The majority of physiological experiments examining auditory nerve behaviour have adhered to the same basic format: a sound stimulus is applied to the cochlea—typically a sinusoid—

Class	spikes per second	% of sample
low-spontaneous	0.5 or fewer	16%
mid-spontaneous	0.5 – 18	23%
high-spontaneous	18 or more	61%

Table 2.1: Classification of auditory nerve fibres in the cat according to spontaneous rate (Liberman, 1978).

and a microelectrode placed in contact with a nerve fibre records the time of each action potential (Moore, 2004).

The average firing rate of an auditory nerve cell has been measured in response to sinusoids of various frequencies and intensities. The first thing to note is that nerve cells fire spontaneously in the absence of any stimulus. Liberman (1978) measured the spontaneous firing rate of 738 nerve cells in the cat and by inspecting a histogram of the sample, was able to identify three distinct groups of cell, characterised by a low, medium and high spontaneous firing rate. His results are summarised in Table 2.1. Spontaneous activity is evident in the left-hand plot of Figure 2.3 at very low sound pressures, particularly in the case of `MCL94-225`.

Cell Thresholds and Neural Tuning Curves

The *threshold* of an auditory nerve fibre is a measure of its sensitivity to a stimulus. Kiang and Moxon (1974) define the threshold as the sound pressure level, in decibels, required to increase the firing rate of the cell above the spontaneous rate by ten spikes per second. The *tuning curve* for an auditory neuron is obtained by measuring the cell's threshold in response to sinusoidal stimuli over a range of frequencies. The frequency associated with the lowest threshold—the frequency to which the cell is most responsive—is termed the *characteristic frequency*. Examples of tuning curves for auditory nerve fibres with characteristic frequencies of approximately 1 kHz are shown in the right-hand plot of Figure 2.3.

The profile of a fibre tuning curve around its characteristic frequency resembles the inverted magnitude response of a band-pass filter. The shape of the tuning curve derives from the mechanical tuning of the basilar membrane at the place where the fibre is sited, provided that active cochlear mechanisms are in effect. The threshold of a neuron at its characteristic frequency is related to its spontaneous firing rate, as reported by Liberman (1978) and more recently confirmed by Yates (1991). Specifically, fibres with a low threshold tend to be associated with a high spontaneous firing rate and *vice versa* (*cf.* Figure 2.3). From Table 2.1, one may infer that the majority of auditory nerve cells in the cat possess a low threshold.

Saturation and Adaptation

The sensitivity of a fibre to the particular stimulus frequencies was likened above to the effect of a band-pass filter upon the amplitude of a sinusoid. This analogy must be

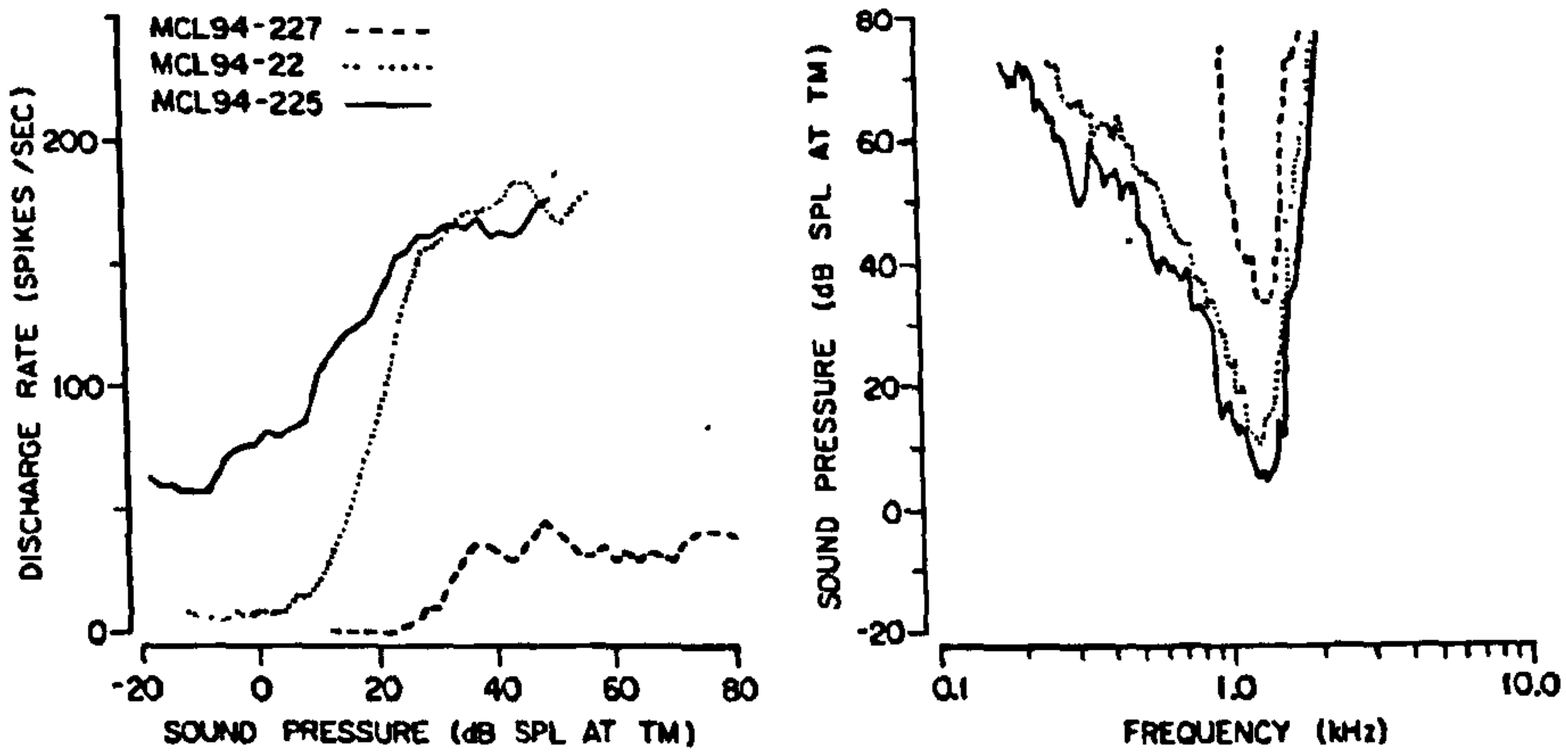


Figure 2.3: Left: experimental rate-level curves showing the average spiking rate for a low- (dashed), medium- (dotted) and high-spontaneous rate fibre (solid), as a function of stimulus level. Right: neural tuning curves for the same three cells. Reused with permission from M. Charles Liberman, The Journal of the Acoustical Society of America, 63, 442 (1978). Copyright 1978, Acoustical Society of America.

applied with some care, as the relationship between the stimulus intensity and average firing rate is neither linear nor time-invariant.

Studies examining how the average rate varies with the intensity of a sinusoidal stimulus have revealed a sigmoidal (rather than linear) relationship (Moore, 2004). This relationship is plotted as a *rate-versus-level* curve: at very low stimulus intensities, the curve is a constant, as the fibre is unresponsive and fires at its spontaneous rate; the mid-portion of the curve is monotonically increasing and relates an increase in intensity to an increase in spiking rate; beyond a certain sound intensity, the cell becomes *saturated*, and increases in intensity do not elicit any further increase in activity. Experimental rate-versus-level curves are shown in the left-hand plot of Figure 2.3. `MCL94-22` is a cell which fires spontaneously below ~10 dB and is saturated above ~30 dB.

The average firing rate of an auditory fibre, rather than being an instantaneous function of the stimulus frequency and intensity, displays a degree of adaptation over time. A helpful way to characterise this dynamic behaviour is a *post-stimulus time histogram* (PSTH), which divides the time immediately following the application of a stimulus into short time intervals, or 'bins', and counts how many discharges occur during each one. Figure 2.4 shows a summary PSTH recorded from a fibre following a 180 ms tone burst at its characteristic frequency (Kiang, 1980). This PSTH is typical of most auditory nerve fibres: the introduction of a tonal stimulus produces an elevated firing rate, which gradually decays towards an *adapted rate*; when the stimulus is released, the cell activity returns to the spontaneous rate.

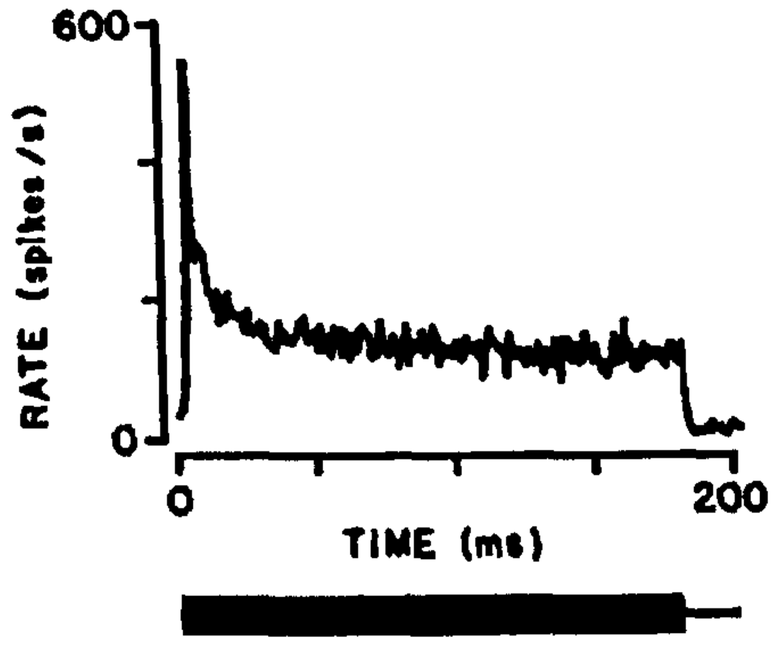


Figure 2.4: A post-stimulus time histogram for an auditory nerve cell in the cat. Reused with permission from Nelson Y. S. Kiang, The Journal of the Acoustical Society of America, 68, 830 (1980). Copyright 1980, Acoustical Society of America.

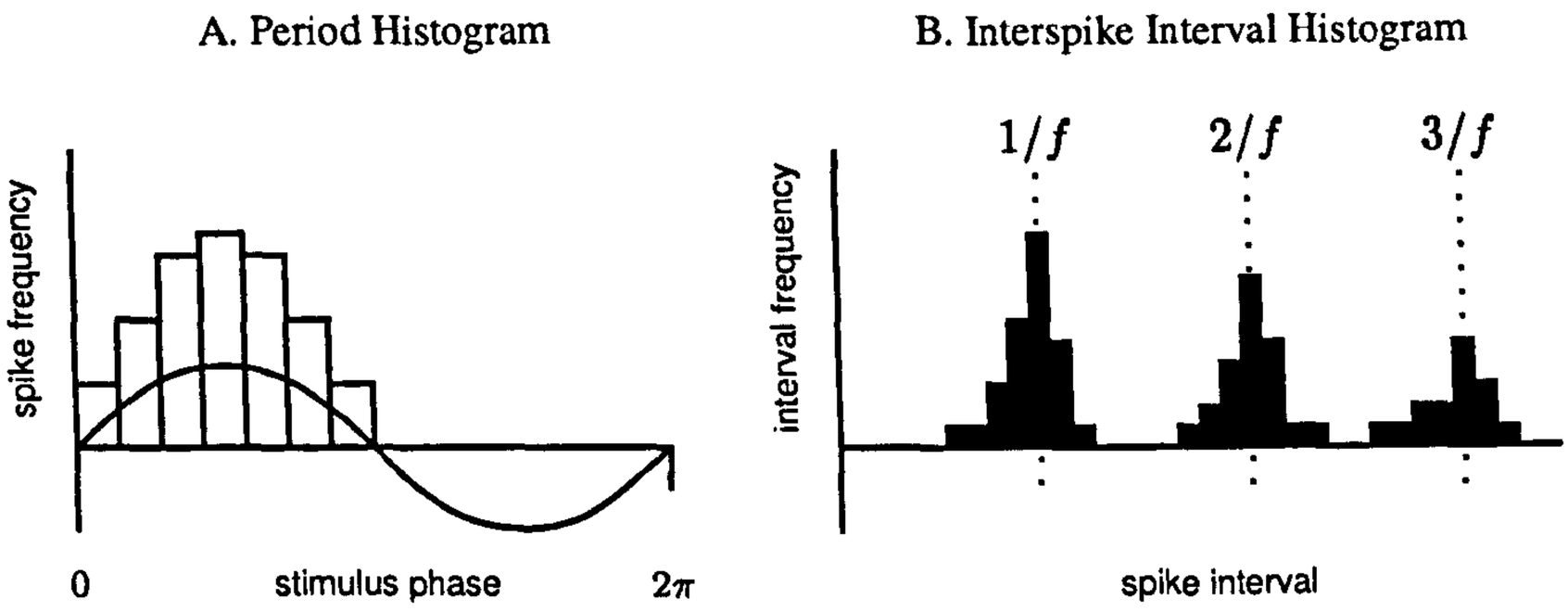


Figure 2.5: A) an illustrative period histogram showing the preference of an auditory nerve cell to fire in phase with the stimulus. B) a typical interspike interval histogram. See Pickles (1988, page 90).

Phase Locking

The features of a sound stimulus are represented not only in the average firing rate of an auditory fibre but also in the timing of individual discharges—the *fine structure* of the response. An auditory nerve fibre generates spikes randomly; however, the distribution of spikes in time is governed to a some extent by the stimulating waveform. Specifically, the nerve activity evoked in response to a tone is typically concentrated at a particular phase, in which case the fibre is said to be *phase-locked*. This section only considers the activity evoked in a single fibre by a tone; however, phase-locking is also observed in response to sub-threshold tones, two-tone stimuli (Brugge et al., 1969), AM broadband noise envelope, fundamental frequency (Javel, 1980) and even complex signals, such as speech.

The presence and degree of phase-locking in an auditory nerve fibre is revealed by a *period histogram*, a graph which relates how many spikes occur at each phase of a tone

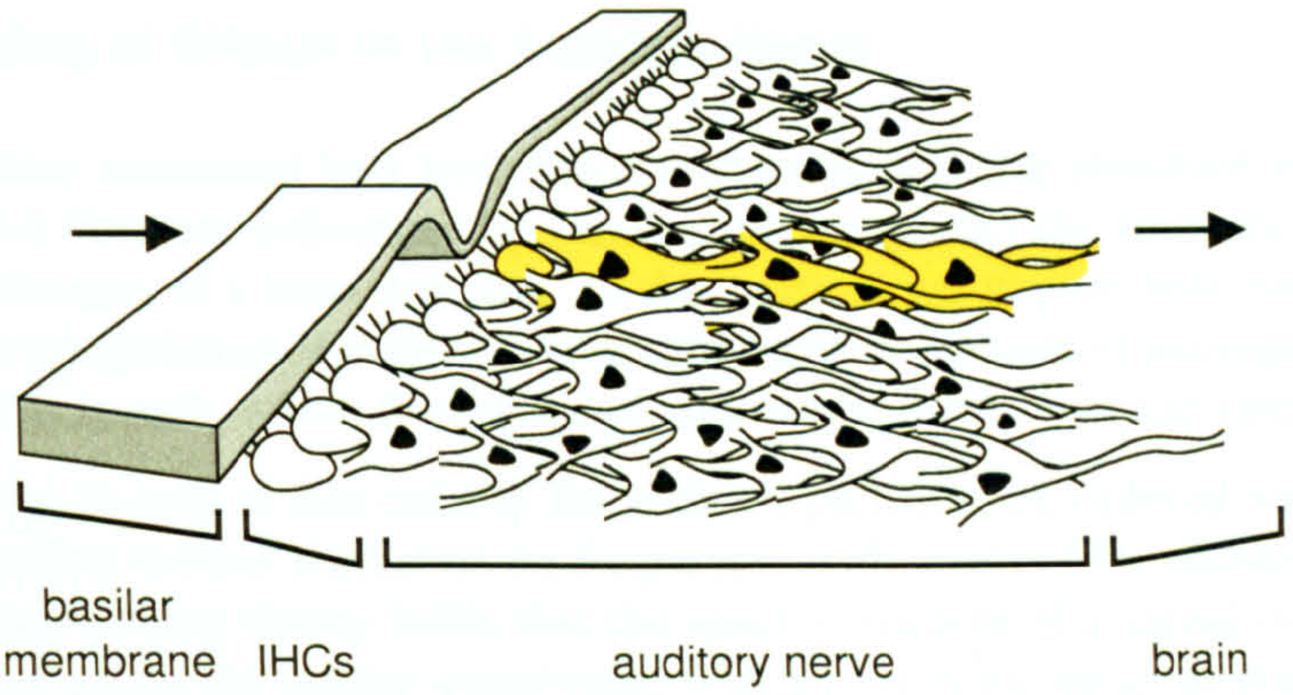


Figure 2.6: A tone stimulus produces a peak displacement on the basilar membrane. The neighbouring inner hair cells transduce this motion into a neural signal, which is preserved tonotopically throughout the auditory pathway.

stimulus. Figure 2.5A shows a period histogram—phase-locked in this case to $\pi/2$—typical of those obtained in physiological experiments. Spikes are observed almost exclusively during one half-period of the stimulus, during which time the probability of a spike is related to the waveform. Consequently, in a time-averaged sense, the pattern of discharges proceeding from a fibre conveys a half-wave rectified version of the basilar membrane motion.

A second means of characterising the fine structure of a fibre's response is to measure the timing of spikes in relation to each other, rather than in relation to the stimulus. One common approach is to form a histogram from the intervals between consecutive spikes, such as the one depicted in Figure 2.5B. The results conform to that which one might expect, given that spikes are generated intermittently and are phase-locked when they do occur. If a tone with frequency f Hz is presented, and a spike occurs in every cycle, then the interspike intervals will exhibit some variability around $1/f$ seconds—the first mode in Figure 2.5B. If spikes occur on alternate cycles, then the intervals will vary around $2/f$ seconds and so on.

It is generally held that the natural variability in the discharge times of fibres leads to the deterioration of phase-locking at frequencies above around 4–5 kHz in most mammals (e.g., Palmer and Russell, 1986). This can be explained in terms of Figure 2.5B. If the tone frequency, f, is increased, but the variability in timing remains constant, then the peaks of the interval histogram merge together; consequently, the strong periodic component in the histogram is lost. (See the *synchronisation index* of Young and Sachs (1979) for one measure of the strength of phase-locking.) The interspike interval histogram of a cell is also influenced by its *refractory period*, i.e., the duration following a discharge in which the cell is recovering and cannot fire. If the refractory period exceeds N/f, the first N peaks of the interval histogram are absent.

2.1.4 The Encoding of Stimuli in the Auditory Nerve

So far we have examined how both the average rate and fine structure in the firing of an isolated fibre are influenced by (and so might encode) the frequency, intensity and phase changes of a tone stimulus. In this section, we inquire how the pattern of excitation in a population of auditory nerve cells conveys features of a complex acoustic signal, such as speech, which the brain and, ultimately, the listener can interpret.

The first thing to note is that cells in the auditory pathway are ordered *tonotopically*, that is, according to their characteristic frequency, at all levels of the auditory pathway. The *rate-place coding* theory holds that the spectral content of a signal is manifested as resonances along the basilar membrane, which in turn excite local populations of auditory neurons; as a result, the spectral envelope of a stimulus is preserved by the average firing rate in a cross-section of the auditory nerve. There is a substantial body of experimental evidence that supports the rate-place hypothesis. In studies of the cat, for example, Sachs and Young (1979) have shown that the formants of vowel sounds presented at low intensities are apparent in measurements of the average-rate profile.

The rate-place theory is sufficient to explain how low-intensity sounds are represented in the auditory nerve; however, for stimulus intensities beyond 40 dB—a moderate sound intensity, equivalent to the level of quiet conversation—the majority of nerve cells are saturated (see Section 2.1.3 above) and the formant peaks are no longer discernable (Greenberg and Ainsworth, 2006). Furthermore, the *two-tone suppression* phenomenon, in which the stronger of two tones diminishes the response of the weaker (Moore, 2004; Sachs and Kiang, 1968), would seem to argue against the place encoding of a tone complex at high intensities. Current formulations of the rate-place theory rely upon the broader dynamic range of low-spontaneous fibres to preserve spectral features at higher sound intensities (Sachs et al., 2006).

The apparent shortcomings of the rate-place theory have prompted some researchers to seek a complementary (or alternative) explanation, based on the fine timing of nerve discharges. These *temporal coding* theories maintain that phase-locking in auditory nerve fibres is exploited in frequency analysis, e.g., in the intervals between discharges. It is generally accepted that timing information in a neural signal degrades as it is transmitted via synapses along the auditory pathway. The trend, as one moves from the periphery to the central processing areas, is that cells exhibiting phase-locking are relatively fewer and can only encode lower frequencies (Meyer, 2006).

Despite the lack of empirical evidence for widespread phase-locking in the higher centres of the auditory pathway, researchers have proposed several mechanisms by which temporal features may be preserved. First, the *volley principle*, first proposed by Wever (1949), suggests that greater precision in phase-locking may be achieved by summing the response over a population of fibres, each of which contributes phase-locked spikes intermittently. Second, it is possible (but not proven) that temporal features are encoded as an average rate signal at some early stage of the auditory processing chain. Interspike intervals are well-represented in the PSTH response of primary-like cells in the cochlear nucleus (Sachs et al., 1988); the auditory midbrain is also a candidate for this conversion process (Meyer, 2006). A recent remark from Møller (2006, page 115) suggests

> "[the] assumption that phase-locking of neural discharges deteriorates in synaptic transmission may be incorrect and the need to convert the temporal code into a spike rate code or a spatial code is not as urgent as earlier assumed."

The apparent phase-locking limitations of cells in the auditory nerve does not then imply that fine time structure is lost thereafter. On the contrary, data recorded from bushy cells located in cochlear nucleus has demonstrated the *improvement* of temporal resolution following the spatial integration of the neural signal over many input synapses (Joris et al., 1994).

In summary, the extent to which the auditory system exploits spectral and temporal information is still unclear and remains the subject of ongoing experimentation and debate. Many researchers now support a duplex theory, which proposes a representation incorporating place-code information and fine temporal detail. To this effect, Møller (2006, page 118) writes: "Contemporary research indicates that both place and temporal coding are important for frequency discrimination in the auditory nervous system."

2.1.5 Auditory Scene Analysis

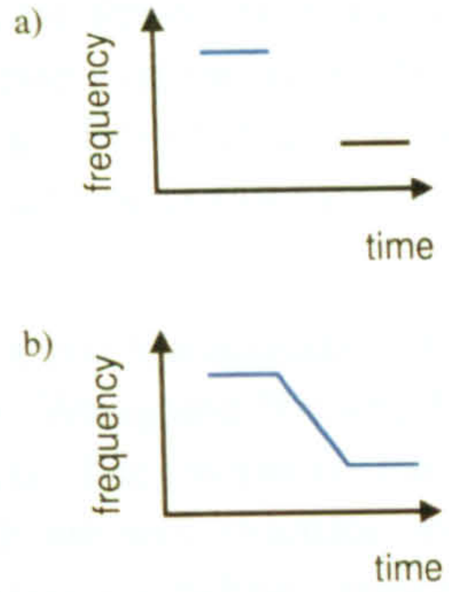
The sections above have sketched some theories concerning the way in which basic features of an acoustic signal—time, frequency and intensity—might be encoded in the auditory nerve. However, when listening in everyday mode (Gaver, 1993), we are principally aware not of signal properties, nor even their psychophysical correlates (loudness, pitch, etc.), but rather *whole entities*. For instance, when we hear the word "car" spoken, our impression is that of a single perceptual unit, rather than a complex tone preceded by a noise burst. (We shall ignore the issue of *analytical listening*, in which the deliberate focus of attention is the acoustic signal itself.) This supports the view of Gestalt psychologists, who argue for an innate, biological tendency to arrange a perceptual field, or raw sensory data, into whole objects (Koffka, 1935).

Remarkably, the auditory system not only combines disparate acoustic features into a unified, auditory experience, but is also able to distinguish individual sources of sound within a mixture. This is a surprising fact considering that, in many contexts, sound reaching the ears contains energy contributed by different sources, which overlaps in time and frequency. Bregman (1990, page 2) provides the following example:

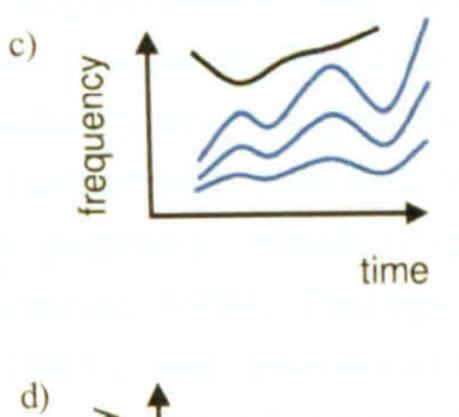
> "A friend's voice has the same perceived timbre in a quiet room as at a cocktail party. Yet at the party, the set of frequency components arising from that voice is mixed at the listener's ear with frequency components from other sources. The total spectrum of energy that reaches the ear may be quite different in different environments. To recognise the unique timbre of the voice we have to isolate the frequency components that are responsible for it from others that are present at the same time. A wrong choice of frequency components would change the perceived timbre of the voice. The fact that we can usually recognise the timbre implies that we regularly choose the right components in different contexts."

similarity This principle says that elements with similar properties should be grouped. These properties may include a common fundamental frequency, spatial location, timbre, or modulation in envelope or frequency.

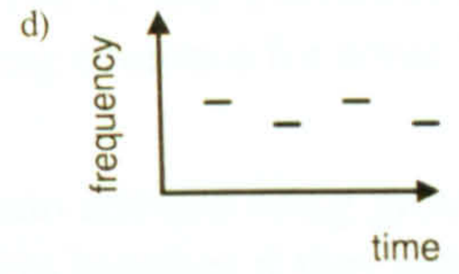
good continuity Smooth variation promotes the perception of a unified, changing sound. For instance, if two tones are presented separately, as shown in (a), then they are perceived individually; if the tones are connected by a *glissando*, as shown in (b), then a single, dynamic sound is perceived (Bregman and Dannenbring, 1973).



common fate This principle states that components which vary identically in some property should be grouped together. Of the four frequency components sketched in (c), three form a single stream on the basis of a common changes in fundamental frequency, and the fourth is heard separately.



proximity This principle refers to the tendency for elements which are close in time or frequency to be perceptually fused. The alternating tone bursts depicted in (d) are heard as a single stream. When the tone bursts are spaced closely in time, i.e., played rapidly, the low and high frequency tones form separate streams, as shown in (e). See van Noorden (1975).



closure The closure principle groups elements if they appear to be fragments of a continuous element which has been masked by noise. For example, (f) shows three tonal sweeps which are isolated from each other by silent gaps. If these gaps are filled with noise bursts sufficiently intense to act as a masker, the closure principle prefers a long, modulated tone obscured by occasional transient interruptions, as (g) illustrates (Ciocca and Bregman, 1987).

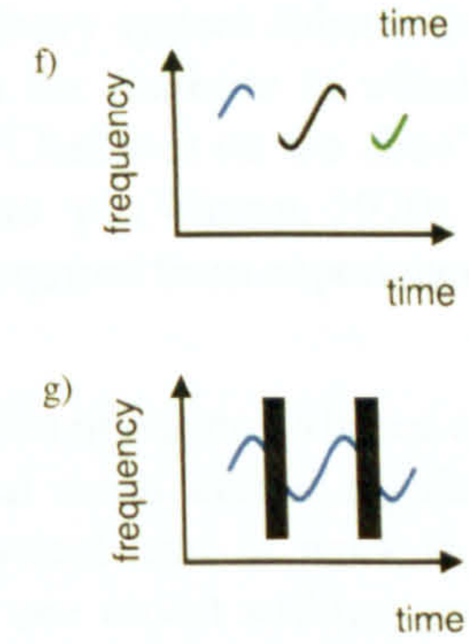


Table 2.2: Five primitive grouping cues.

The task the auditory system performs in "choosing the right components in different contexts" is called *auditory scene analysis* (ASA) (Bregman, 1990). When discussing sound, Bregman draws an important distinction at the outset between *sources* and *streams*. For example, when we say "the sound of a violin", to what are we referring? The term 'sound' may be employed in a mechanical sense—the vibration of the strings and the surrounding air when the bow is drawn across—in which case the violin is a *source*, and each note is an *acoustic event*. Alternatively, 'sound' may relate to a listener's mental impression of the violin sound, in which case the notes of the violin blend together into a single *stream*. Streams may also aggregate to form larger streams, in hierarchical fashion, depending on the mode of listening. For example, a violin, cello, flute and oboe may be attended as four individual streams or as one stream, e.g., a quartet.

Bregman presents auditory scene analysis as a two-stage process: the acoustic signal is analysed into a collection of sensory elements, or *segments* (Wang and Brown, 2006), then perceptual streams are synthesised from disjoint subsets of segments on the basis of either top-down or primitive grouping principles, which we will examine shortly. Physiological accounts of either stage—presuming the analysis-synthesis concept of ASA is correct—are lacking in detail at present, although the segmentation stage would clearly require an auditory mechanism to identify and extract prominent features in a signal. The temporal-tonotopic encoding of a signal in the auditory nerve suggests that such a process might operate in the time-frequency domain; furthermore, the discovery of specialised neurons in the higher centres of the auditory pathway, which respond selectively to amplitude and frequency modulation (Eggermont, 1994; Phillips and Hall, 1987), onsets or offsets (e.g., Whitfield and Evans, 1965), and interaural time difference (Brugge and Merzenich, 1973), provide compelling evidence for some kind of 'feature extraction' layer.

The synthesis stage in ASA forms collections of segments into streams using grouping cues. *Top-down* or *schema-driven* grouping cues bind features together if they match a learned perceptual pattern. In other words, top-down cues are drawn from a listener's previous experience of sound in order to assess which segments of a sound scene belong together. The influence of top-down processing in the auditory system is demonstrated by studies of the *phoneme restoration effect*. This phenomenon is observed when a short section of a word is replaced with noise, and the auditory system substitutes the noise for an alternative, meaningful sound, depending on the sentence in which the word is embedded. For example, the noise bursts (□) in "□eel was on the shoe" and "□eel was on the orange" are respectively heard as 'h' and 'p' (Warren, 1970). The top-down cues operating in these experiments are clearly acquired from experience and may differ from person to person.

Primitive or *data-driven* grouping cues refer to the association of segments based on the Gestalt principles of perceptual organisation. In the natural world, certain regularities are impressed upon sounds as a consequence of the physical laws at work in their production. For instance, many sounds are generated by one object striking another, e.g., tapping a pencil on a table top. From an acoustic perspective, such actions result in the *simultaneous* introduction of sound energy across many frequencies, giving rise to the principle: "sounds that share a common onset should be grouped". Another major

class of mechanical system encountered in nature, besides transient disturbances, is made up of oscillatory systems, which produce sounds by rapidly repeating the same action. Periodic sounds always possess a harmonic spectrum; thus, we obtain a second principle: "partials that share a common fundamental should be grouped". The Gestalt principles appear to be aligned with these tendencies and are considered innate rather than learned. Table 2.2 lists five abstract grouping principles for which experimental evidence exists. (See Bregman (1990) for a review.)

2.1.6 Interim Summary

The hearing process begins with pressure variations at the ear, which propagate along the auditory canal and cause the membranous eardrum to vibrate. These vibrations are then communicated through three small bones (ossicles) onto another, smaller membrane called the oval window, which forms an opening at the base of the cochlea. The difference in size between the eardrum and oval window, along with the configuration of the ossicles, helps to overcome the impedance step encountered by a wave travelling from air to fluid.

The motion of the oval window causes longitudinal waves in the fluid-filled chambers of the cochlea, which in turn excite traverse travelling waves in the basilar membrane. The mechanical properties of the basilar membrane vary along its length so that high frequencies produce a peak displacement at the base (near the oval window) and low frequencies produce a peak displacement near the apex. The local displacement of the basilar membrane results in the deflection of the hairs (stereocilia) attached to inner hair cells in that region.

With each deflection of a hair, a small amount of neurotransmitter is released into the space between the inner hair cell and an auditory nerve fibre. When a sufficient quantity of transmitter has accumulated in the cleft, the fibre generates an action potential, which in turn causes a cascade of spikes to flow along the afferent pathway from the auditory periphery to the brain. There remains some disagreement concerning which mechanisms are responsible for the encoding of a complex stimulus in the auditory nerve. The rate-place theory holds that the average firing rate in a tonotopic array of cells provides a time-varying spectral representation. Temporal (and duplex) theories assign a role to the timing of cell discharges in frequency encoding, especially for intense stimuli, which saturate the average-rate response.

The rate-place and timing theories of encoding broadly account for a listener's ability to interpret sounds presented in isolation; however, the physiological basis for the ability to organise a mixture of sounds into separate perceptual entities is less well-understood. Auditory scene analysis is a conceptual model which supposes that the auditory system segregates a signal into time-frequency segments and then reassembles those segments into streams according to a set of grouping cues. Top-down cues are established from experience and assign segments to the same stream if they resemble a learned pattern (or 'scheme'). Primitive cues associate segments according to Gestalt principles, such as similarity, proximity, closure, continuity and common fate.

2.2 Computational Models

The remainder of this chapter is devoted to the subject of *computational auditory models*, namely, algorithms that are designed to mimic either the biological function of the ear, the behaviour of the listener, or both. The task of modelling the entire hearing process is usually presented as a matter of identifying functional blocks within the ear—these we described in the first half of the chapter, e.g., the outer-middle ear, the cochlea, the auditory nerve and so forth—along with their inputs and outputs, producing a computer program to simulate the behaviour of each one at a suitable level of abstraction, and then assembling the programs into a chain. The first goal of this section is to review the various auditory models that have been proposed for the following three major functional stages: the outer and middle ear (§2.2.1), the basilar membrane (§2.2.2) and neuro-mechanical transduction (§2.2.3).

The second part of the section explores the various ways in which researchers have joined together the component auditory models to form complete systems, in particular, systems that emphasise the role of timing mechanisms in auditory processing. Five categories of temporal processing model are identified in the auditory modelling literature, which we shall enumerate in Section 2.2.4. The first three categories are presented as models of stimulus encoding in the auditory nerve (§2.2.4), and the final two are discussed in connection with computational auditory scene analysis (§2.2.5).

2.2.1 Modelling the Outer and Middle Ear

The combined effect of the outer and middle ear can be modelled by a linear filter that provides a broad resonance around 2.5 kHz and perhaps a second, lesser resonance at 5.5 kHz due to the external ear (Pickles, 1988). The equal loudness contour provided by ISO suggests a transfer function which boosts frequencies in the 2–4 kHz range (ISO, 2003). Other researchers use a high-pass pre-emphasis filter, based on a transfer function originally measured by Lynch et al. (1982) in the cat, modified to match data obtained from humans (Kates, 1991; Slaney, 1988). In a more recent auditory model, Lopez-Poveda and Meddis (2001) combine two FIR filters in series, one which characterises the headphone pressure-to-eardrum pressure system, and the other the eardrum pressure-to-stapes velocity system. The impulse responses for both filters are obtained by applying an inverse discrete Fourier transform to empirical frequency responses (Pralong and Carlile, 1996; Goode et al., 1994).

Unlike some components of the auditory system, the twin purposes of the outer and middle ear—(i) to increase the sound pressure at the oval window and (ii) to modify the incoming spectrum to pre-emphasise the band occupied by speech signals and to assist directional hearing (Moore, 2004)—appear to be highly system- (physiology) and signal- (speech) specific. This particular stage will therefore be excluded from any generalised model of auditory-*style* processing.

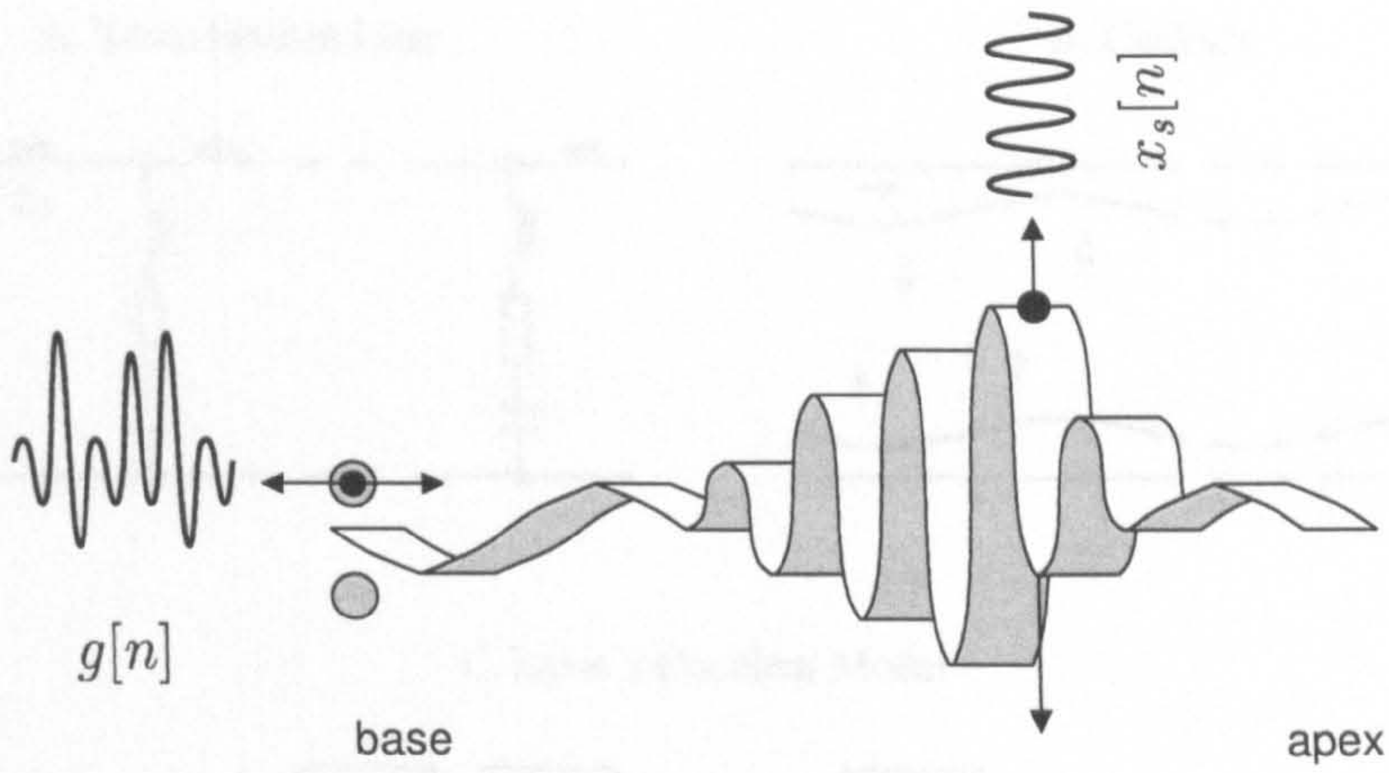


Figure 2.7: The goal is to reproduce computationally the system which transforms a time-varying pressure at the oval window, $g[n]$, into the motion of a single point on the basilar membrane, $x_s[n]$.

2.2.2 Modelling the Basilar Membrane

The vibration at a given place on the basilar membrane in response to a stimulus derives from the physical properties of the membrane, primarily its stiffness and mass, and any active tuning mediated by the outer hair cells. An appropriate model for this behaviour is a digital filter, which transforms a stimulus signal, $g[n]$, into a signal representing the motion of the basilar membrane, $x_s[n]$, at a place indexed by s, as illustrated in Figure 2.7. The output of a bank of filters, i.e., $x_1[n], x_2[n], \ldots, x_M[n]$, then models the motion of the basilar membrane at discrete points along its length.

Transmission Lines

The earliest computer simulations of the motion of travelling waves along the basilar membrane employed the *transmission line* model (Zwislocki, 1948). The velocity of the stapes is modelled by the current density at the leftmost end of the transmission line, and the pressure differences between the scalae are modelled by the voltages across the terminal rails. The line is divided into small elements consisting of the RLC circuit depicted in Figure 1.1.1. The elemental inductance and resistance are assumed to be uniform throughout the transmission line; however, the capacitance grows exponentially with distance from the voltage source. In accordance with the analogy, the mass and friction of the basilar membrane are taken to be uniform, whilst its compliance increases with distance from the oval window (Allen, 1985).

In these models, successive sections of the transmission line act like notch filters. The AC current enters each section, and the flow divides between the parallel and series branches in proportion to the admittance encountered at each: if the parallel branch resonates at the AC frequency, then charge preferentially flows there; if not, the charge flows along the series branch. Returning once again to the physical case, when pressure

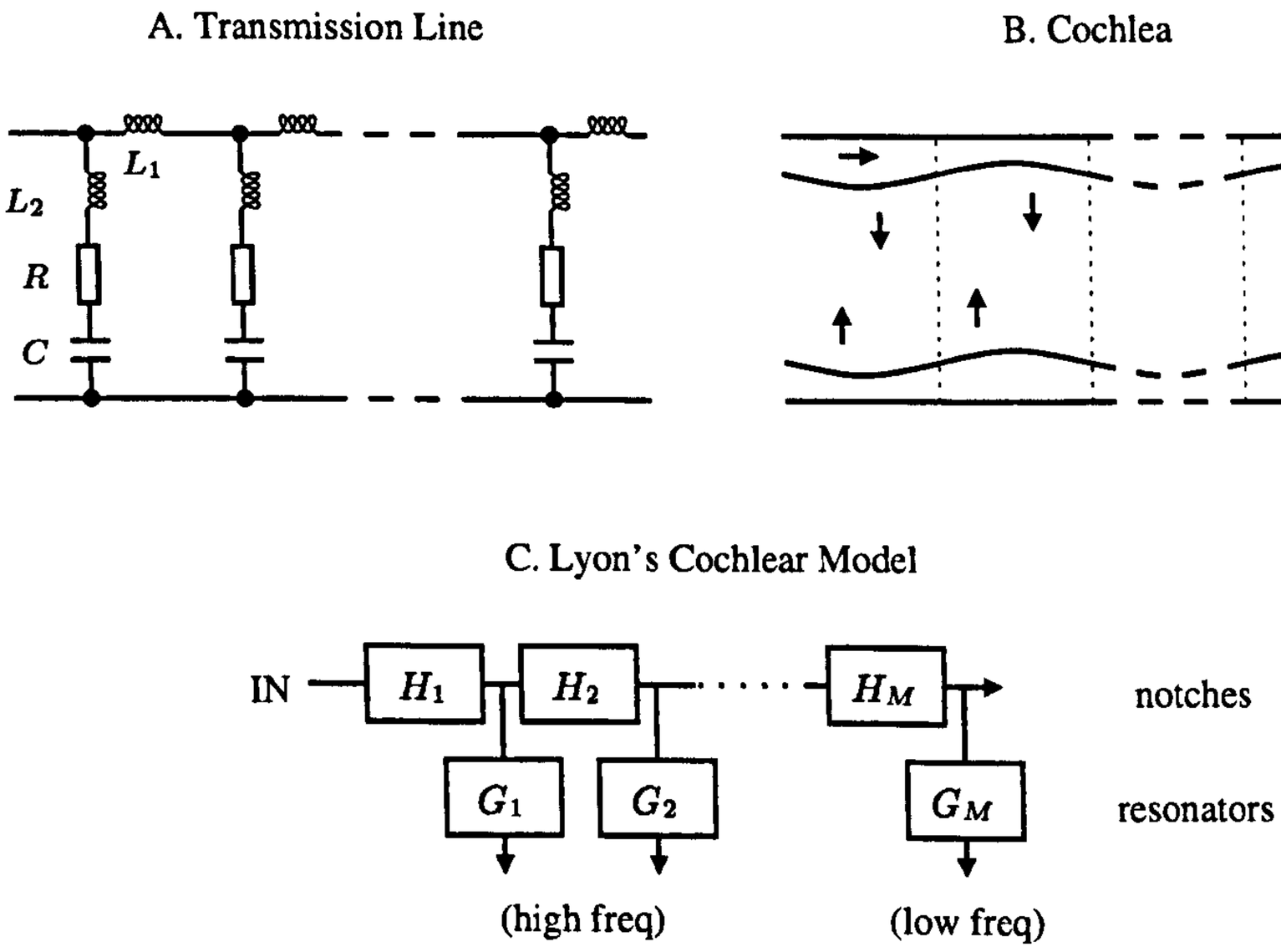


Figure 2.8: Transmission line model of the cochlea. A) the transmission line is composed of a cascade of parallel sections. The inductances L_1 and L_2 model inertia in the fluid and basilar membrane, respectively. The resistance R models loss of energy due to friction. The capacitance C models the compliance of the basilar membrane; the C's increase exponentially along the transmission line (Zwislocki, 1948). B) simplified view of the mechanical forces in the cochlea. The downward facing arrows represent the forces due to inertia, the upward arrows represent the restoring forces due to stiffness, and the right-facing arrows denote fluid flow. C) A schematic showing the stages of filtering in a transmission line model (Lyon, 1982).

is applied to the fluid it can either push against the basilar membrane and store mass in the displacement, or it can propagate along the chamber. If the inertial forces (due to mass) and the restoring forces (due to stiffness) are balanced in such a way that the section moves back and forth with the flow, then the pressure wave directed along the scala diminishes.

The computational model of vibration in the cochlea presented by Lyon (1982), based on the earlier work of Schroeder (1973) and Zweig et al. (1976), utilises a cascade of notch filters to progressively remove the high frequency content from the fluid pressure wave; then, at each stage, a parallel 'resonance' filter converts the pressure on the basilar membrane into displacement. A block diagram representing this process is shown in Figure 2.8C. The displacement of the basilar membrane is read off from the taps along the bottom of the transmission line. The centre frequencies of the sections descend in equal jumps on a quasi-logarithmic scale (Slaney, 1988), and there is a constant-Q relationship between bandwidth and frequency. The magnitude response for eight sections of the transmission line model are shown in Figure 2.9A. Note that,

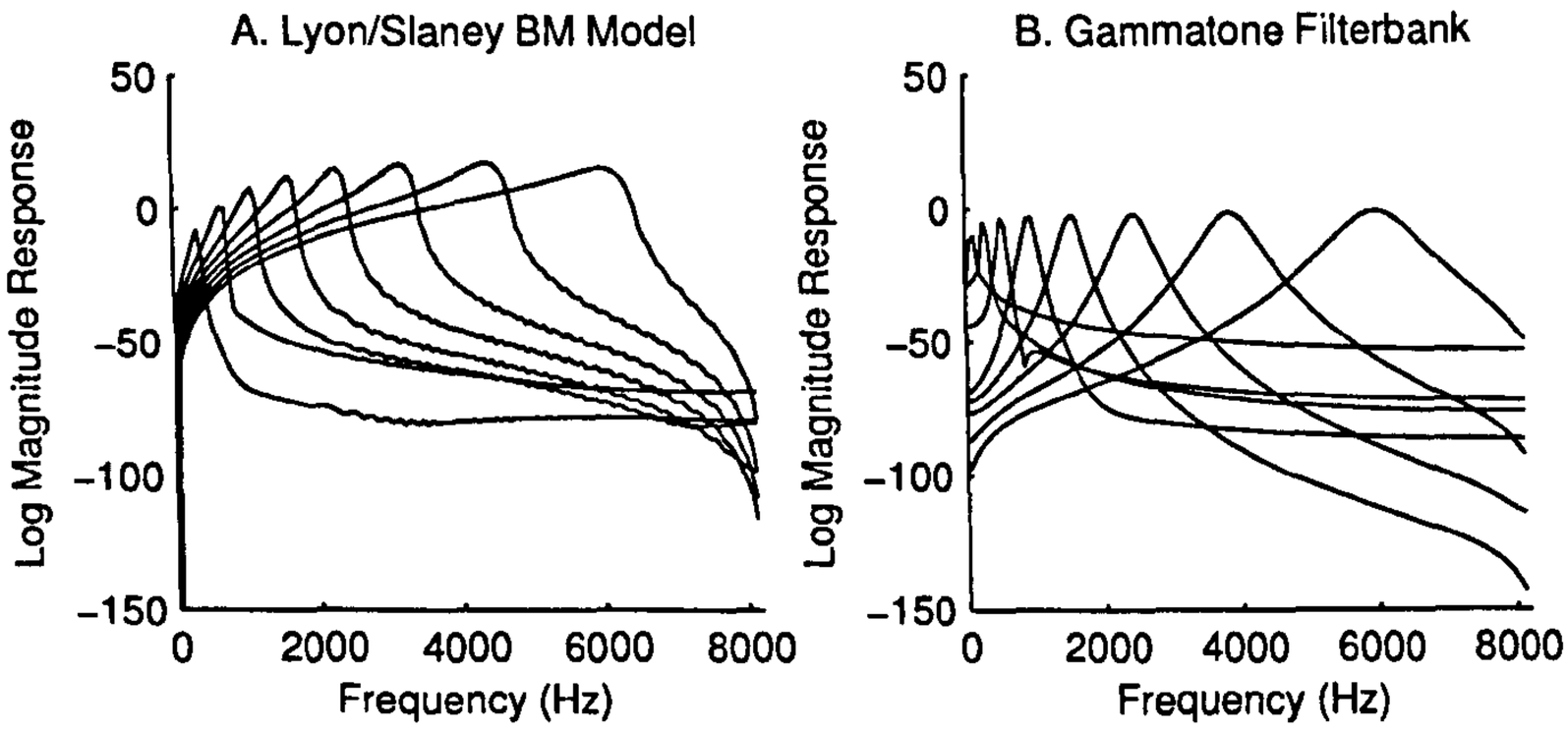


Figure 2.9: Magnitude response of two basilar membrane models. A) eight filters from a transmission line basilar membrane model, produced using code from Malcolm Slaney's MATLAB Auditory Toolbox (Slaney, 1994). B) eight gammatone filters uniformly spaced on an ERB scale between 100 Hz and 6000 Hz.

due to the cumulative low-pass filtering of the serial notch filters, the low-frequency filters exhibit a steep right-hand roll-off.

Filterbank Models

The second type of approach to characterising the response of the basilar membrane to a stimulus is a *system identification* approach. First, the system that transforms the input signal (the motion of the stapes) into the output signal (the vibration of a point on the basilar membrane) is assumed to be linear and time-invariant. Second, a standard experimental procedure is applied to obtain the impulse response of the linear system by measuring its response to, e.g., a click or white noise. A third, optional stage seeks an elegant analytical expression for the impulse response, in which the place on the basilar membrane is a parameter.

The first two steps outlined above were carried out by de Boer and de Jongh (1978), who used a *reverse correlation* technique to discover the impulse responses relating an input stimulus to the firing rate of auditory neurons at various places on the basilar membrane of the cat. Shortly afterward, de Boer (1979, cited by Schofield) completed the third step, publishing a parametric version of the impulse response, called the 'revcor function', which agreed closely with those obtained empirically. Nowadays, this function is referred to as the *gammatone function* and has the form

$$h_{gt}(t) = t^{n-1} \exp(-2\pi B t) \cos 2\pi f_c t, \quad \text{for } t \geq 0, \tag{2.1}$$

where the parameters n, B and f_c refer to filter order, bandwidth and centre frequency, respectively. The magnitude response corresponding to the impulse response (2.1) with order $n = 4$, provides a good fit to psychoacoustic auditory filter shapes (Schofield, 1985).

The transfer function of the nth-order gammatone filter can be obtained from (2.1) using some standard properties of the unilateral Laplace transform:

$$\mathscr{H}_{gt}(s) = \frac{(n-1)!}{2}\left[\frac{1}{(s+2\pi(B-if_c))^n} + \frac{1}{(s+2\pi(B+if_c))^n}\right]. \tag{2.2}$$

For a discussion of the implementation of the gammatone filter in discrete time, see Cooke (1991/1993, Appendix A).

The gammatone impulse response specifies the response at one place on the basilar membrane; a bank of gammatones is needed to model the basilar membrane in its entirety. To construct a filterbank, a principled way to choose the bandwidths (B) and centre frequencies (f_c) of the constituent filters must be found. The bandwidth of auditory filters can be measured using a notched-noise technique (Patterson, 1976). The equivalent rectangular bandwidth (ERB) (Moore, 2004) of an auditory filter has been found to share an approximately linear relationship with its centre frequency (Glasberg and Moore, 1990) and is given by the *ERB function*[1]:

$$B = \mathrm{ERB}(f_c) = 24.7(4.37 f_c \cdot 10^{-3} + 1), \text{Hz}. \tag{2.3}$$

The ERB function enables us to assign a bandwidth to all the filters in the gammatone filterbank, provided that their centre frequencies are known. The centre frequencies themselves can be chosen arbitrarily, e.g., distributed evenly on a linear scale. A better approach, however, is to space the centres evenly on an *ERB scale* (Moore, 2004), with the desirable consequence that the spacing between filters grows at the same rate as their bandwidths change. A frequency f_c has a value on the ERB scale given by the formula (Glasberg and Moore, 1990)

$$\text{ERB number}\,(f_c) = \frac{(\ln 10)\log_{10}(4.37 f_c \cdot 10^{-3} + 1)}{24.7 \times 4.37 \cdot 10^{-3}}, \tag{2.4}$$

which is derived from (2.3). Figure 2.9B plots the magnitude response of a filterbank containing eight gammatone filters. Some notable features of the gammatone filter include its linearity, its relatively broad magnitude response which is symmetric[2] around the centre frequency f_c, and the inexpensive implementation it offers in the form of a digital IIR filter with only $2n-1$ coefficients (Cooke, 1991/1993).

The first two properties of the gammatone filter named above, linearity and passband symmetry, are difficult to reconcile with empirical measurements of the basilar membrane, which reveal a level-dependent, asymmetric frequency response. Several alternative filterbanks have been developed to address these shortcomings. Two specific examples worth mentioning here because of their close relation to the gammatone filter are the *dual-resonance nonlinear* (DRNL) filter (Lopez-Poveda and Meddis, 2001) and the *gammachirp* filter (Irino and Patterson, 1997). Both simulate the broadening of auditory filter bandwidth that results from increasing the stimulus level (Glasberg and Moore, 2000).

[1] For $n = 4$, a correction factor of 1.019 is often applied to (2.3), following Patterson et al. (1988).

[2] Strictly speaking, the general gammatone filter is *asymmetric* around f_c. In an auditory filterbank, $B \approx cf_c$, where c is sufficiently small that the magnitude response may be considered symmetric around its peak for all practical purposes.

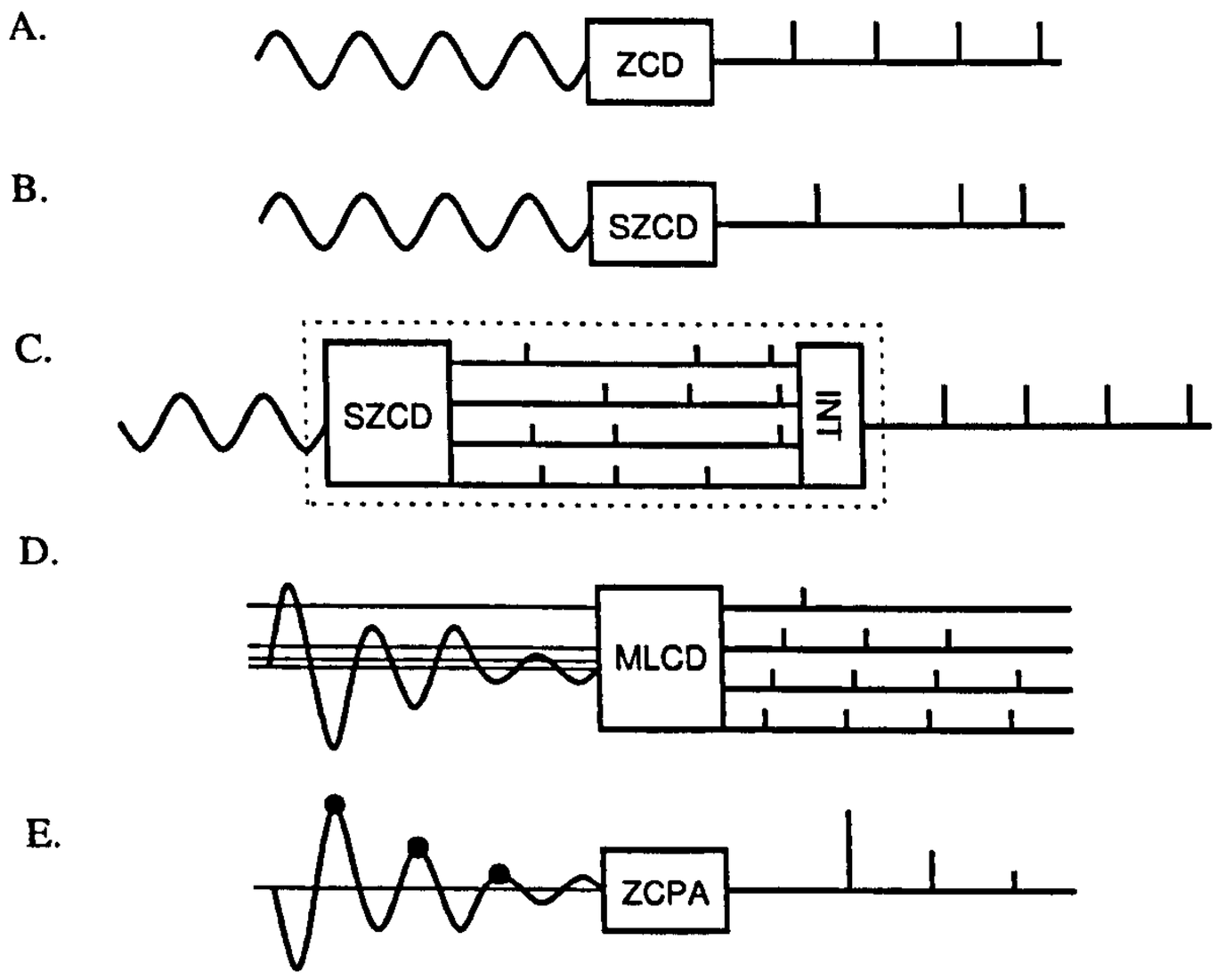


Figure 2.10: Zero crossing transduction models. A) deterministic zero crossing detector (ZCD); B) stochastic zero crossing detector (SZCD); discharges are noisy and intermittent; C) the output of several SZCDs is fed to an integrator (INT), which averages the crossing times of nearly-coincident spikes; D) multi-level crossing detector (MLCD); E) zero crossings weighted by the peak amplitude (ZCPA).

2.2.3 Modelling Neuro-mechanical Transduction

The next block in the auditory periphery model represents the transduction process, that is, the conversion of a mechanical signal to a neural signal by the inner hair cells. The displacement of points along the basilar membrane is supplied as input, and the output is a measure of nerve activity across an array of model inner hair cells. Depending on the implementation, the output of a single IHC model may be a spike train (a point process describing the timing of model discharges), a spike rate or a spike probability.

Zero Crossing or Level Crossing Detector

One simple approach to modelling the auditory-neural transduction process is a zero crossing detector, which generates a spike whenever the signal changes sign from negative to positive. This choice of model effectively discards information concerning the extent to which the basilar membrane is displaced and retains only the frequency (and phase) of its vibration. This is a suitable model insofar as it emphasises the preservation of timing in auditory nerve fibres; in this respect, the zero crossing detector may be considered the 'ideal phase-locked unit'. Furthermore, as the majority of fibres are saturated at moderate stimulus levels, the zero crossing detector also represents the 'ideal saturated unit' (Figure 2.10A).

The fidelity of the zero crossing model may be questioned on at least two grounds. First, the record of basilar membrane motion present in the spike train of an IHC is probabilistic and incomplete; a zero crossing detector, in contrast, is deterministic and generates a spike on every cycle. One possible response to this criticism is to introduce a stochastic element into the zero crossing detection process so that i) a spike only results from a crossing with probability p, and ii) a spike time is computed by adding a small amount of noise to the crossing time. This method yields a model spike train, which is less artificial in appearance (Figure 2.10B). Alternatively, the zero crossing detector can be said to simulate the behaviour of fibres that integrate the response of a population of noisy cells—the volley theory (Figure 2.10C).

A second, more serious objection to the zero crossing model is that a population of zero crossing detectors does not explicitly encode the envelope of the signal. In fact, all auditory neurons respond across some dynamic range; low-spontaneous rate cells, though in the minority, remain responsive even at high stimulus levels. One variant upon the zero crossing model uses a *multi-level crossing detector* (Ghitza, 1988) to encode both timing and amplitude information. This scheme encodes up-going level crossings, and the output of the unit is a spike train for each level (Figure 2.10D). A practical alternative to using multiple levels is to weight each spike by a function of peak amplitude across the previous interval; this is equivalent to a 'continuum' of levels (Kim et al., 1999) (Figure 2.10E). The application of these models is discussed in greater detail in the next section.

Half-wave Rectification, Compression and Automatic Gain Control

The appeal of choosing a multi-level or weighted zero crossing detector lies in their ability to capture basic auditory transduction phenomena—phase-locking, level compression and saturation—using only very simple components. Another popular class of functional model combines half-wave rectification, compression and automatic gain control (AGC) to achieve a closer match to physiological data. For instance, the cochlear model of Lyon (1982) performs a half-wave rectification in each channel, followed by a three-stage adaptation process implemented by a series of automatic gain controls. The time constants on each AGC differ and account for various sources of adaptation in the ear. In Lyon's model, the compressive nonlinearity is included as the final stage, although its effect is incorporated into the AGC feedback.

Seneff's model of the auditory periphery includes a similar hair cell transduction stage accomplished by a half-wave rectification, compression and AGC (Seneff, 1988). The half-wave rectification and compression take the form of a static nonlinearity; the adaptation stage is performed by two nonlinear time-invariant filters. The in-channel transduction model of Dau et al. (1996) consists of a half-wave rectifier and low-pass filter, followed by an adaptation stage implemented as a cascade of automatic gain controls with time constants ranging between 5 ms and 500 ms.

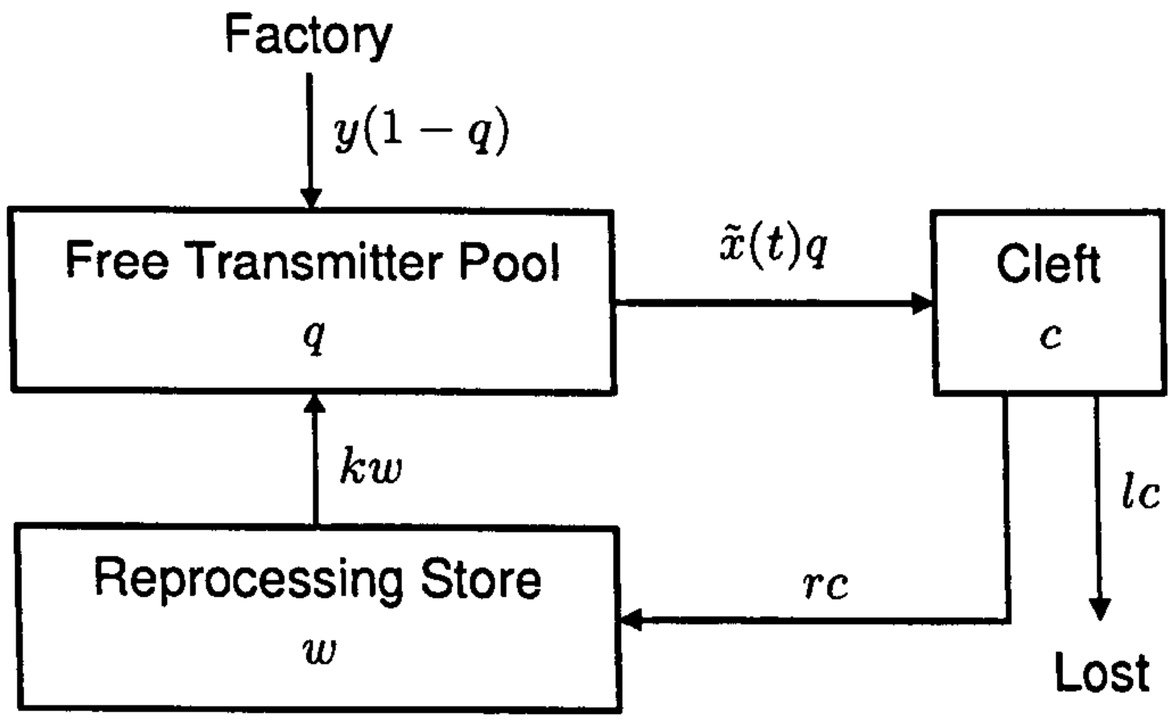


Figure 2.11: A diagram depicting the flow of transmitter between the three reservoirs of the Meddis model, along with a source (factory) and sink (loss), adapted from Meddis (1986, 1988).

Meddis' Hair Cell Model

A widely-adopted model of the inner hair cell is that of Meddis (1986)[1]. Like Seneff's model, the Meddis hair cell (MHC) is able to reproduce a variety of characteristics, including phase-locking, compression, saturation, adaptation and spontaneous firing (Meddis, 1988). The MHC simulates the manufacture, transfer, recycling and loss of chemical transmitter substance within, and in the vicinity of, the inner hair cell, and specifies three 'reservoirs' in which a quantity of transmitter may reside: the free transmitter pool, the cleft and the reprocessing store. These quantities are respectively represented by the state variables q, c and w, and differential equations describe how transmitter is exchanged between them. The physical understanding of these equations is described next and schematically presented in Figure 2.11.

The free transmitter pool leaks into the cleft in proportion to the compressed, half-wave rectified motion of the basilar membrane, $\tilde{x}(t)$. At the same time, the free transmitter pool is replenished in proportion y to how empty it is (where $q = 1$ is considered full), and a proportion k of substance in the reprocessing store is also recovered. This gives the first differential equation governing flow in and out of the free pool:

$$\frac{dq}{dt} = -\tilde{x}(t)q(t) + y(1-q(t)) + kw(t). \tag{2.5}$$

The fluid in the cleft arrives exclusively from the free transmitter pool, as described above, and from here, a proportion l is lost, and a proportion r is taken up into the reprocessing store. The reprocessing store simply returns transmitter fluid from the cleft to the free transmitter pool. (The rate at which fluid enters and exits reprocessing depends on the choice of r and k.) From these statements follow two further differential

[1] For a more recent revision of this model, see Sumner et al. (2002).

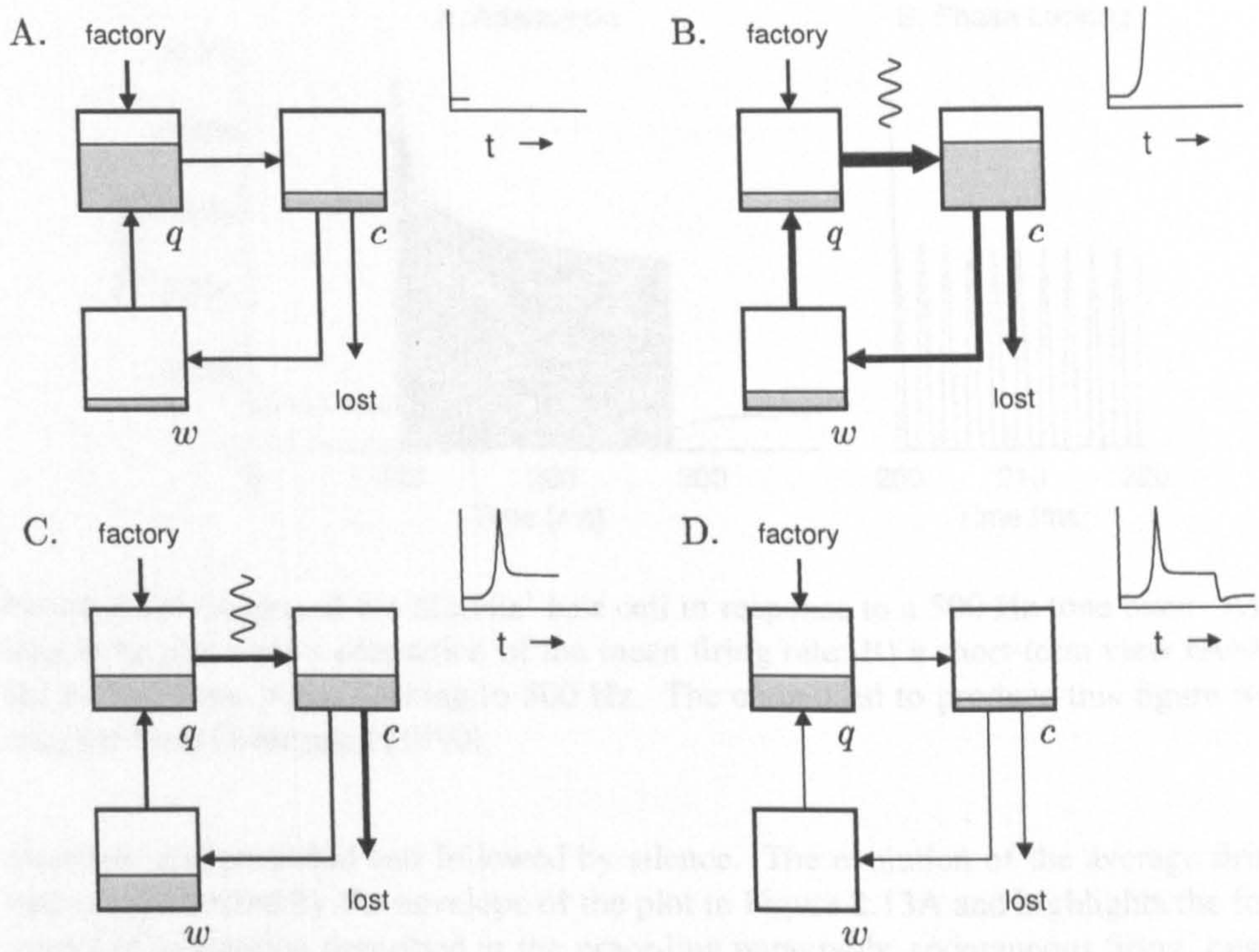


Figure 2.12: An illustration of four qualitative stages of adaptation in the Meddis hair cell model for a medium-intensity tone burst: A) no stimulus, spontaneous activity; B) onset rate; C) adapted rate; D) offset and recovery of spontaneous activity.

equations:

$$\frac{dc}{dt} = \tilde{x}(t)q(t) - lc(t) - rc(t) \tag{2.6}$$

$$\frac{dw}{dt} = rc(t) - kw(t). \tag{2.7}$$

The output of the MHC is a spike probability, directly proportional to the amount of transmitter in the cleft, c. The operation of the model can be understood as follows. Prior to the application of a stimulus, the free transmitter pool is almost full, and a small amount of transmitter fluid passes into the cleft, triggering spontaneous firing (Figure 2.12A). When the stimulus is initially presented, the free transmitter floods into the cleft, producing a sudden increase in firing probability (Figure 2.12B). Gradually, the cleft content is depleted through loss and reuptake, and the factory can only replenish it at a limited rate, via the free transmitter pool. This gives rise to the adapted rate (Figure 2.12C). When the stimulus is removed, spontaneous firing resumes at a reduced rate, as the free transmitter pool is now empty (Figure 2.12D). Once the factory has restored the free transmitter pool, the process can begin again.

A MATLAB implementation of the MHC was used to produce Figure 2.13, which illustrates two aspects of the model's response to a 500 Hz tone burst, 180 ms in

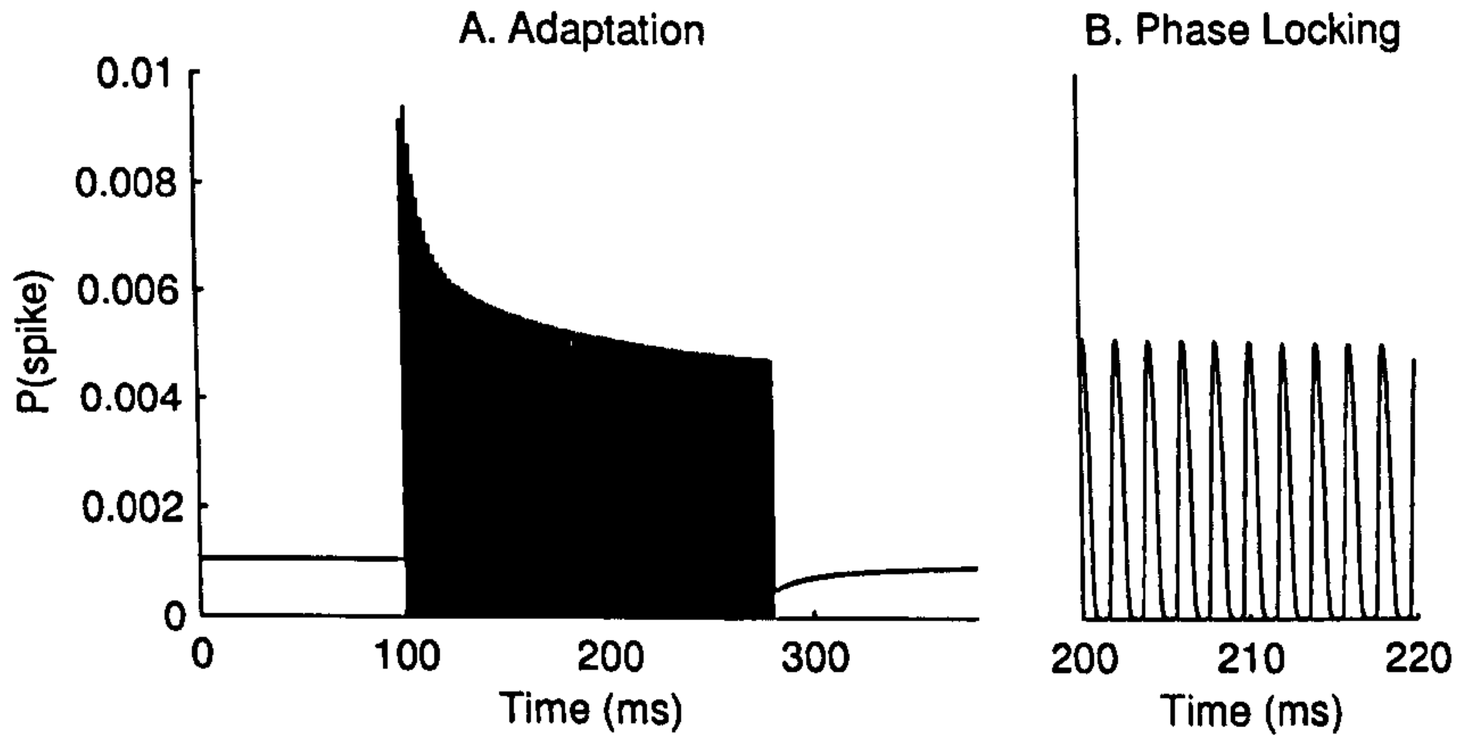


Figure 2.13: Output of the Meddis' hair cell in response to a 500 Hz tone burst. A) a long-term plot shows adaptation of the mean firing rate; B) a short-term view reveals the probabilistic phase-locking to 500 Hz. The code used to produce this figure was adapted from Ostergaard (1990).

duration, and preceded and followed by silence. The evolution of the average firing rate is represented by the envelope of the plot in Figure 2.13A and highlights the four stages of adaptation described in the preceding paragraph: spontaneous firing, onset, adaptation, and reduced spontaneity and recovery. The plot also closely resembles the experimental data of Kiang (1980), reproduced in Figure 2.4. Figure 2.13B magnifies a portion of the left-hand plot, to reveal how the fine structure of the firing probability is phase-locked to the tone frequency; specifically, there are ten peaks equally spaced over a twenty-millisecond period, indicative of a 500 Hz stimulus.

2.2.4 Models of Stimulus Encoding in the Auditory Nerve

Computer models of the auditory periphery simulate the sequence of transformations a signal undergoes from the moment it arrives at the ear as an acoustic wave to the point of its encoding in the auditory nerve as series of spikes. Section 2.1.4 reviewed some theories advanced to explain how the auditory-neural encoding relates to the original sound stimulus. Rate-place theories hold that sounds are encoded by the average firing rate of fibres associated with different places along the basilar membrane, whilst temporal theories emphasise the importance of precise timing preserved by individual nerve spikes. Five approaches to temporal processing are discussed over the next two sections, which we may briefly characterise in advance.

1. *joint synchrony* – a model that represents the extent to which each of its channels is dominated by, or phase-locked to, a periodicity at its centre frequency.

2. *in-channel* – a model that independently extracts frequency information from each of its channels, using the phase or (more specifically) zero crossings, without any reference to centre frequencies. This is sometimes referred to as a *non-place* approach (e.g., Ghitza, 1988).

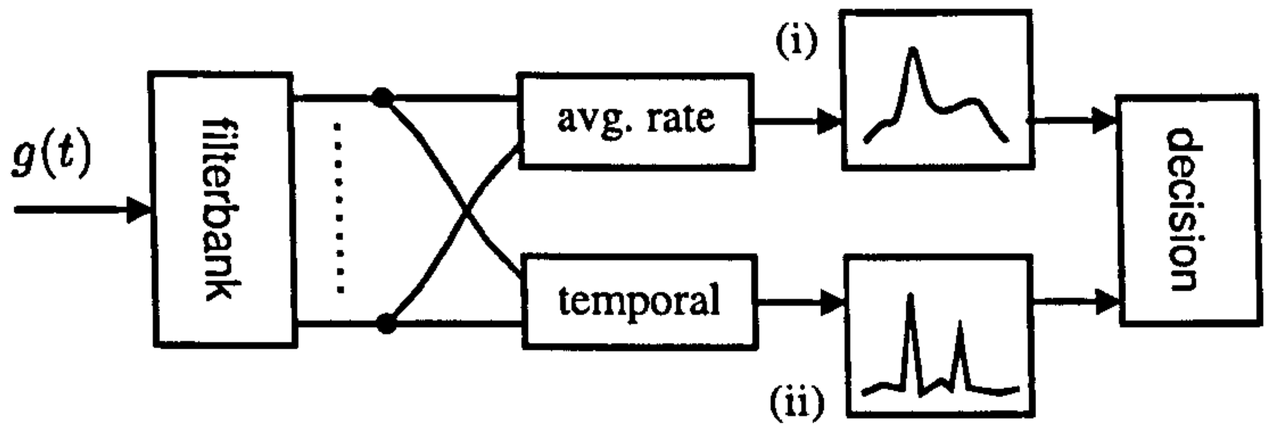


Figure 2.14: Schematic of a joint synchrony / average-rate model. The output of the filterbank follows two independent pathways, which respectively compute (i) an average-rate representation and (ii) a temporal representation. Features from both representations are then supplied to a third stage, e.g., a decision rule or classifier.

3. *spatial-temporal* – a model that extracts frequency or phase-locking information, but does so by comparing the output of many channels. (The term "spatial" does not relate to directional hearing, rather the spatial axis of the basilar membrane.)

4. *synchrony strand* (Cooke, 1991/1993) – a specific model for locating and tracking time-frequency structures using instantaneous frequency information, which is distinct from the three methods above, on account of its parametric output.

5. *correlogram* (Slaney and Lyon, 1990) – a plot that represents timing information by showing the autocorrelation in every channel as a two-dimensional image.

There are without doubt other categories or models that this list omits, and we may assume some degree of overlap amongst the approaches that are listed.

I. *Joint Synchrony / Average-rate Models*

Models that compute two separate representations from the output of a cochlear filterbank, one based on average firing rate, the other based on how closely the fine temporal structure is synchronised to the channel centre frequency, we designate *joint synchrony / average-rate models*[1] and schematise in Figure 2.14.

The auditory model of Seneff (1988) adopts the joint-synchrony / average-rate format. The initial stage of the model is a half-wave rectification, compression and adaptation performed for each output of a forty-channel auditory filterbank. The average-rate spectrum is then calculated by smoothing this representation, whilst the synchrony spectrum is derived, in parallel, by processing the same data with a *generalised synchrony detector* (GSD). The output of the GSD in channel s is given by

$$\mathrm{GSD}_s(t) = \arctan\left(\frac{\langle|x_s(t) + x_s(t-\tau_s)|\rangle}{\langle|x_s(t) - x_s(t-\tau_s)|\rangle}\right), \tag{2.8}$$

[1] following Seneff (1988).

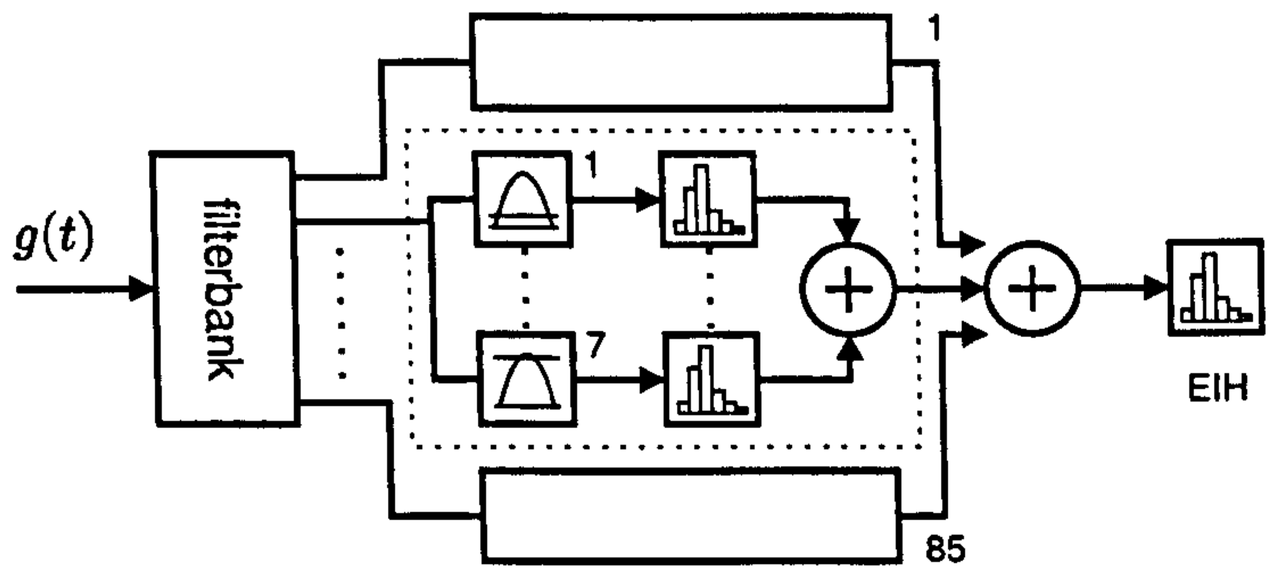


Figure 2.15: A schematic of the EIH. The signal is first decomposed into 85 channels by a cochlear filterbank. Each channel is then processed by 7 level-crossing detectors and the reciprocals of the 20 most recent intervals from each level are compiled into a histogram. The EIH output is the sum of all the intermediate histograms.

where τ_s is a delay corresponding to the reciprocal centre frequency of channel s, and $\langle|\cdot|\rangle$ denotes the time-averaged magnitude. Upon inspecting (2.8), the response of the GSD is seen to be greatest when the input contains a periodicity close to the centre frequency of the channel.

A second auditory model that employs multiple representations is the rate-place and temporal-place encoding scheme proposed by Sachs et al. (1988). Their study bypasses the computational filterbank stage by directly supplying a record of the spike times measured in fibres on the (cat) basilar membrane as input to the algorithm. The rate-place representation plots the normalised average firing rate of each fibre against its characteristic frequency, thus presenting the degree of neural activity on a conventional frequency axis. As the properties of fibres vary considerably, the *normalised average rate* scales the firing rate into the range 0–1, where 0 and 1 respectively correspond to the spontaneous and saturated rate.

The temporal-place representation is computed along a separate pathway and plots the *average localised synchronised rate* (ALSR) for each harmonic of a complex. The extent to which a single fibre is phase-locked to a harmonic is measured by the *synchronisation index*: the magnitude of the Fourier transform of the fibre's PSTH at the harmonic frequency, divided by the average rate. The ALSR for a given harmonic is then defined as the synchronisation index averaged across all fibres. For a more detailed discussion of synchronisation index and ALSR, with examples, see Young and Sachs (1979).

II. (a) *In-channel Temporal Processing: The Ensemble Interval Histogram*

The second class of auditory model reviewed in this section uses the zero crossings in the output of a cochlear filterbank to produce a frequency-domain representation of the signal. The frequency analysis performed by these models is therefore accomplished in two stages: i) each filter passes only a narrow range of frequencies—this is a standard, spectral analysis; ii) the intervals between the zero crossings in the filter output are

used to estimate the precise frequency of local, dominant components. Several auditory models conform to this scheme[1], e.g., Deng and Sheikhzadeh (2006); Fulop and Fitz (2006); Gardner and Magnasco (2006); Kim et al. (1999); Cooke (1991/1993). One of the earliest examples of this kind of auditory model is the *ensemble interval histogram* (EIH), proposed by Ghitza (1988) for noise-robust speech applications. It is instructive to examine the original algorithm in some detail (*cf.* Figure 2.15).

The initial stage of processing in the EIH is a filterbank containing eighty-five cochlear filters, spaced logarithmically between 200 Hz and 3200 Hz, which decomposes the input signal into narrow bands. This simulates the displacement of points on the basilar membrane, as described in Section 2.2.2. The second stage models the transduction process in each channel using seven level crossing detectors, distributed evenly on a log-scale over the dynamic range of the signal (see Section 2.2.3 and Figure 2.10D). Last of all, the EIH representation itself is computed at 5 ms intervals by pooling the reciprocal of the twenty most recent intervals from every level crossing detector into a histogram. The histogram spans the 0–3200 Hz range and is uniformly divided into one hundred bins. The result is a two-dimensional time-frequency representation, in which each histogram 'time slice' is potentially compiled from 11,900 intervals.

The representation of a signal in the EIH can be understood as follows. If a 100 Hz tone, e.g., is presented without noise, then every filter outputs upward level crossings at 10 ms intervals, which accumulate in the 100 Hz histogram bin (Figure 2.16A). How many intervals each filter contributes will depend on the distribution of the levels, the signal amplitude and the attenuation of the filter. A more complicated scenario arises when broadband noise is added to the input signal, so that every filter is driven by a mixture of tone and noise. The level crossing intervals of filters in the vicinity of the tone gravitate towards the tone frequency and produce a peak in the EIH. The remote filters, which are dominated by noise, generate a spread of intervals and together create a relatively flat response in the EIH (Figure 2.16B).

Simple speech sounds are represented in the EIH in a similar way. Vowel sounds are produced when pressure pulses caused by the rapid opening and closing of the glottis excite resonances in the vocal tract (Gold and Morgan, 2000). For modelling purposes, the vocal tract is typically treated as an all-pole linear system. In the frequency domain, the vowel sound is the product of two spectra: a harmonic spectrum (associated with the glottal pulse train), and the response of the vocal tract filter, which consists of a number of resonant peaks or *formants*. Figures 2.17A and 2.17C display the amplitude spectrum for two synthetic vowel sounds: `[ar]` and `[er]`. The corresponding EIH plots are shown in Figures 2.17B and 2.17D. The vowel formants are marked on all the plots, although the fine structure (i.e., individual harmonics) is unresolved in the EIHs at high frequencies. The first formant of both `[ar]` and `[er]` are overresolved in the EIH.

The spectrogram and a vareity of EIHs for a recorded speech signal ("set white with P2 soon", spoken by a male voice) are shown in Figures 2.17E and 2.17F–H, respectively. The original EIH model (Ghitza, 1988) extracted a fixed number of the most recent

[1] Some of these models do not extract component frequencies from zero crossing intervals *per se*, but rely on allied quantities such as inter-peak intervals or the phase-derivative. The principle is identical, however.

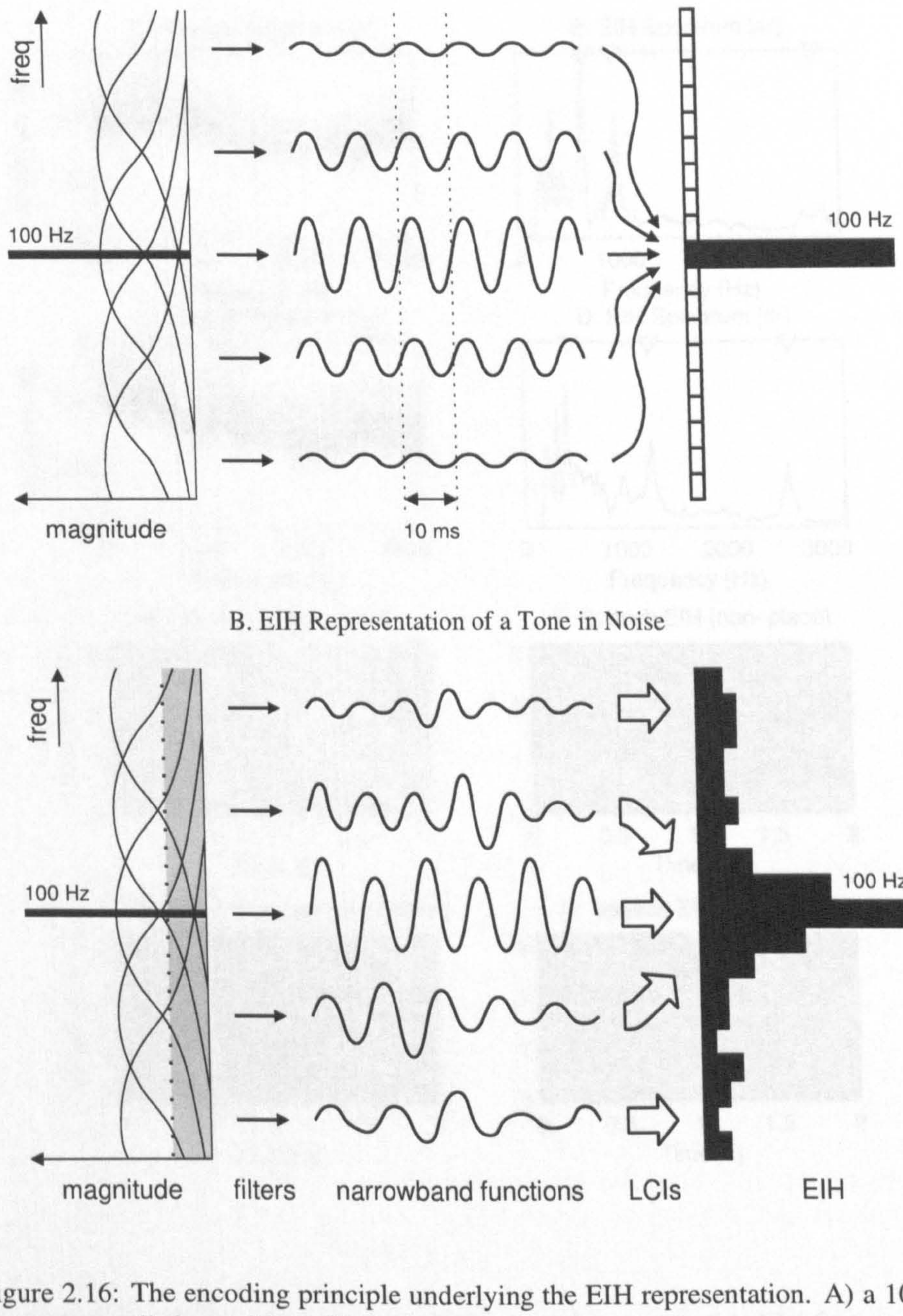


Figure 2.16: The encoding principle underlying the EIH representation. A) a 100 Hz tone causes all the cochlear filters to respond at the tone frequency; every interval contributes to a single frequency bin in the EIH. B) adding noise to the tone generates random output from each filter, except those close to the tone, which gravitate towards the dominant frequency and create a broad peak in the EIH.

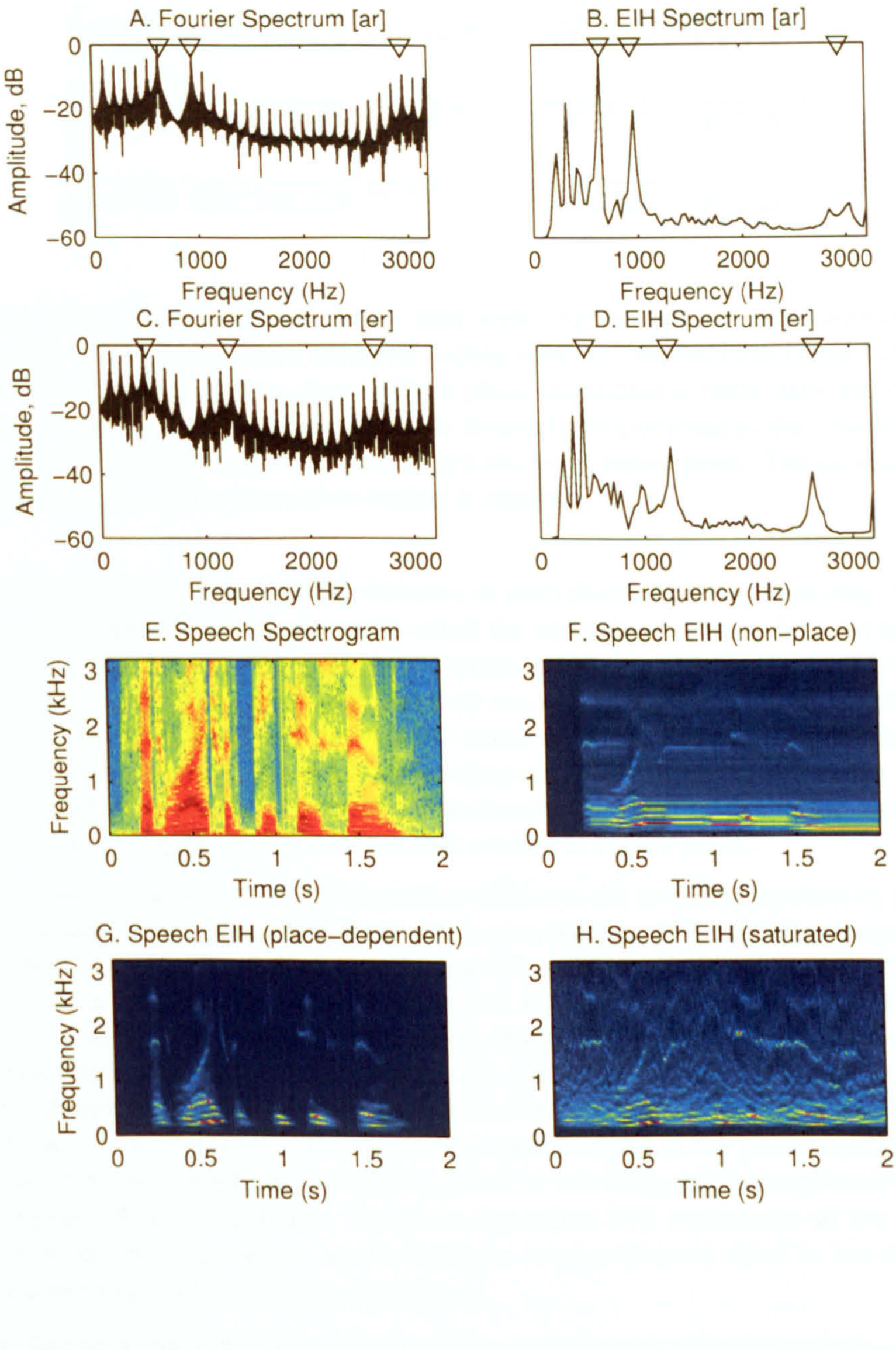


Figure 2.17: A–D) Fourier magnitude spectrum and corresponding EIH for two vowel sounds: `[ar]` and `[er]`. The first three formants are marked on using inverted triangles. E) spectrogram for the utterance "set white with P2 soon" alongside a variety of EIH plots (F–H). See text for details.

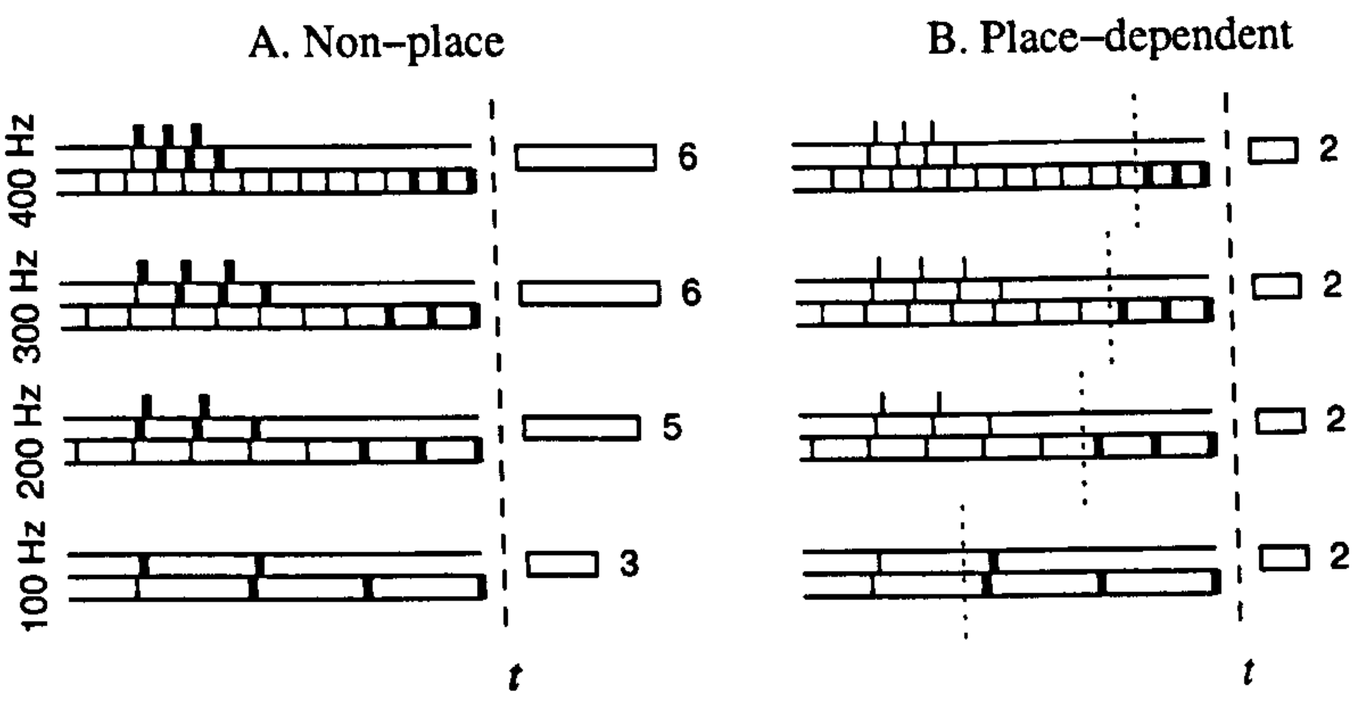


Figure 2.18: Time resolution in an EIH with four filters and three level-crossing detectors. A) in a non-place temporal coding scheme, intervals are formed from the most recent spikes prior to time t. B) a place-dependent scheme uses the intervals within a trailing window, whose length is inversely proportional to the channel centre frequency. The spikes that are counted are shown in heavy print. The localisation in time afforded by place-dependent coding is superior.

intervals from each level crossing detector, in each channel, at each time step. This is a *non-place* approach, as the manner in which the intervals are chosen does not explicitly depend on the channel. One problem with this approach is that level crossing intervals remain the most recent intervals until they are replaced. For instance, a sufficiently high-energy noise burst will activate the entire array of level crossing detectors, and even if silence ensues, the most recent intervals will still contribute to the EIH, as no mechanism is included to cater for their removal. This results in the undesirable 'smearing' along the time axis of the EIH evident in Figure 2.17F.

A subsequent variant of the EIH algorithm addresses the smearing problem by adopting a *place-dependent* approach to interval selection (Ghitza, 1994). In this representation, the EIH is formed from the intervals that fall within a window K/f_c seconds long, where f_c is the channel centre frequency and K is a constant parameter. With respect to time resolution, the non-place and place-dependent EIH behave very similarly when a level crossing detector registers intervals; but when no intervals are registered, the place-dependent version does not use outdated information. This principle is illustrated in Figure 2.18. Figure 2.17G presents the speech signal in the place-dependent EIH. Unlike the non-place EIH, the bright patches in this image are appropriately confined to regions of speech activity. The place-dependent EIH reproduces all the principal features of the spectrogram, whilst highlight some additional detail in low-frequency harmonic structure and formant transitions.

Another problem associated with the EIH—non-place or place-dependent—relates to the distribution of the levels in each channel. The levels used to produce the EIH shown in Figure 2.17G had to be adjusted by trial-and-error before arriving at a reasonable contrast in the image. If the levels are set too high, then the level-crossing detectors are never triggered, and the EIH is empty; alternatively, if the level amplitudes are very small with respect to the output of the cochlear filterbank, then the multi-level crossing

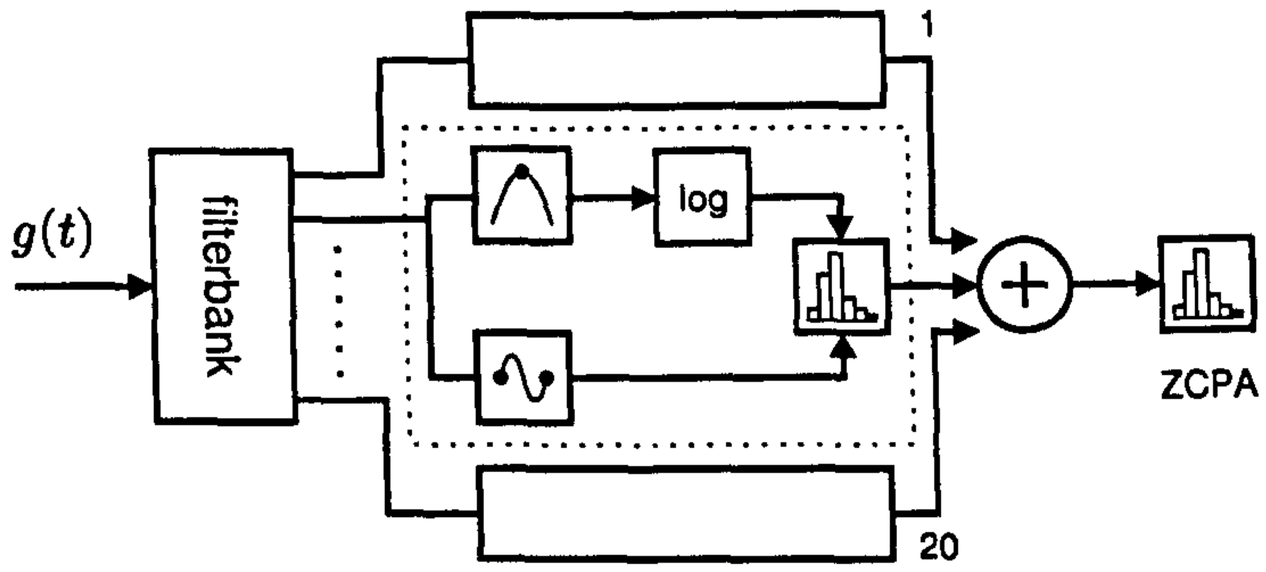


Figure 2.19: A schematic of the ZCPA (*cf.* Fig. 2.15). Each output channel of a cochlear filterbank is processed by a peak detector and zero crossing detector. The contribution of each reciprocal interval to the histogram is weighted by a monotonically-increasing function (e.g., logarithm) of the peak amplitude across that interval.

detector starts to behave like a zero crossing detector. In the latter case, the EIH is 'saturated' and preserves only temporal information: Figure 2.17H shows the saturated EIH for the same speech signal. One can still trace in this representation a surprising amount of formant and harmonic detail, but it is no longer possible to discern energetic and noisy regions.

The EIH parameters specified above were chosen to suit an ASR system operating on telephone speech band-limited between 100 and 3200 Hz and occupying a known dynamic range (Ghitza, 1988). Apart from this context, the EIH is a general-purpose signal processing routine and, with suitable calibration, could conceivably be tailored to other applications, including passive sonar. However, the dependence of the levels upon the signal, demonstrated above and in Figure 2.17, discourages the use of the EIH in contexts where the dynamic range of the signal is unknown.

II. (b) *In-channel Temporal Processing: Zero Crossings with Peak Amplitudes*

A variant of the EIH has been recently proposed by Kim et al. (1999), called the *zero crossings with peak amplitudes* (ZCPA) algorithm. The concept underlying this representation is almost identical to the EIH; however, differences in the way the amplitude is computed avoid at least three faults in the EIH: i) the problem of choosing levels, ii) spurious intervals, and iii) perturbation noise. In this section, we briefly examine how the ZCPA works and then consider each of these problems in turn—and their solution in the ZCPA.

The first stage of the original ZCPA implementation is a bank of twenty cochlear filters (Kates, 1991). The signal in each channel follows two paths: one extracts the zero crossings; the other extracts the peak amplitudes between pairs of zero crossings, which are subsequently compressed using the function

$$G(x) = \log(x + 1). \tag{2.9}$$

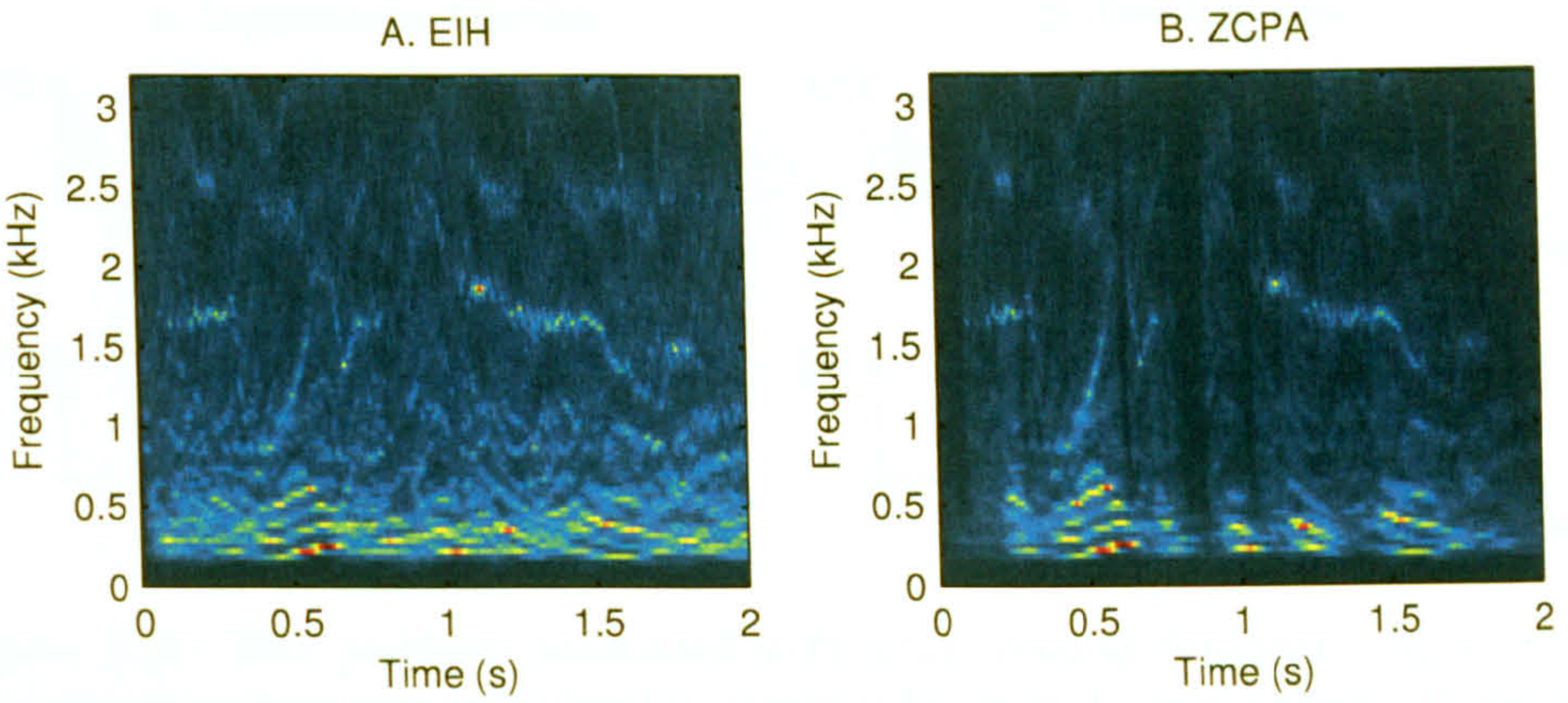


Figure 2.20: A moderately intense speech signal in the A) EIH and B) ZCPA representations. (The speech signal is that used to produce the EIH in Figure 2.17G, increased by a level of 20 dB.) Whilst the EIH has become saturated, energetic regions are still apparent in the ZCPA. Increasing the signal level by a further 10 dB saturates both representations. (These examples use 85 cochlear filters.)

As with the place-dependent EIH, a histogram is formed from the reciprocal zero crossing intervals contained in a trailing window inversely proportional to the channel centre frequency. However, instead of using a multi-level crossing detector to encode the envelope, the contribution that each interval makes to the histogram is determined by a non-linear function of peak amplitude measured across that interval. The ZCPA thus encodes frequency and amplitude information within each channel independently: the former by the zero crossings, the latter by the peak amplitudes. The ZCPA output is a summary histogram found by adding the twenty minor histograms together.

The paragraphs above identified the most serious drawback to using the EIH in a wider context, namely, the difficulty in assigning levels to the level crossing detectors in a principled way. It should be clear that the ZCPA, by relying on peak amplitudes rather than level crossings, is able to represent intensity on a continuous scale ranging from zero to infinity. In one sense, the peak amplitude can be likened to the joint contribution of a 'continuum' of level crossings detectors, spanning the entire positive range; the ZCPA is, in this view, an ideal implementation of the EIH, not constrained by the practical cost of computation associated with a super-dense array of level crossing detectors.

In a biological ear, the majority of auditory nerve fibres are saturated by moderately intense sounds (*cf.* Section 2.1.3). Similarly, saturation, which in general terms refers to amplitude measurements being 'clipped' in some way, is one of the consequences that follows from a poor choice of detector levels in the EIH. In this section, we must ask: Does the ZCPA become saturated as the input level increases? Because the ZCPA encodes amplitude continuously and without upper limit, we might expect that the answer is no; however, because the ZCPA formulated in Kim et al. (1999) incorporates a log-compression, a form of saturation *does* take place. To see why this is the case, imagine that n intervals of signal contributing to the bin k have peaks P_i; then the value

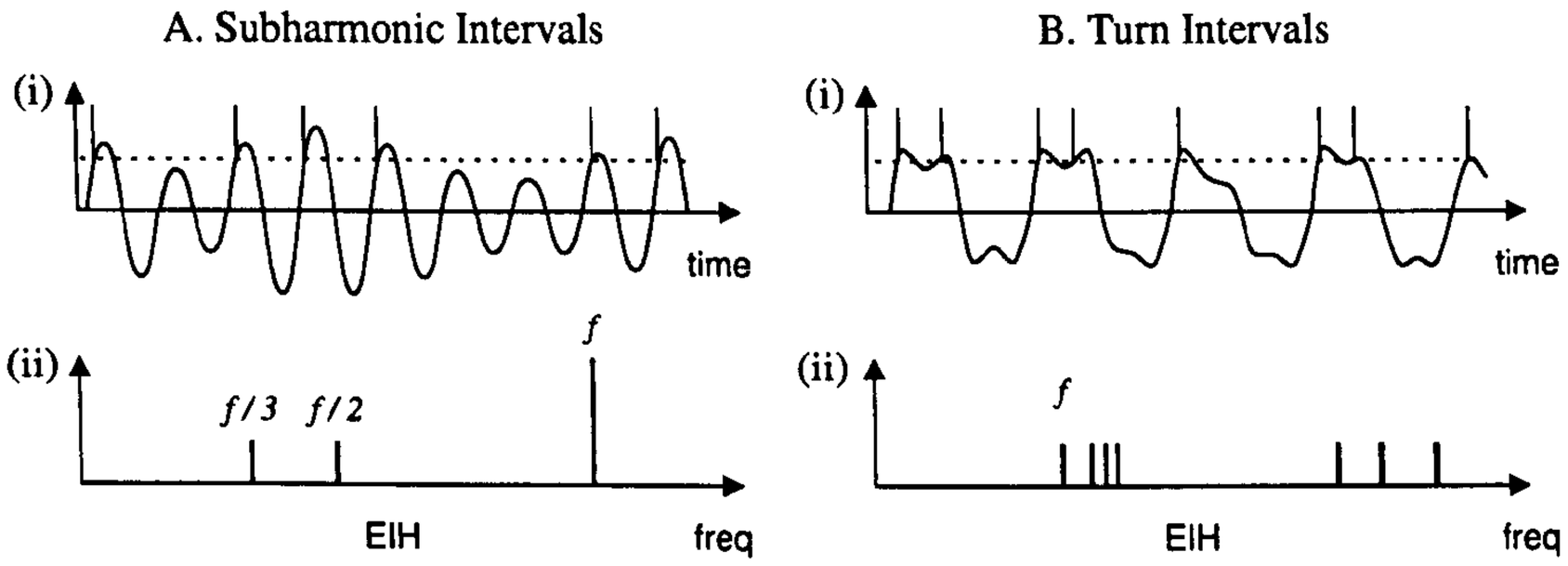


Figure 2.21: Two problems associated with level crossing detectors. A) a gently modulated envelope may cause level crossings to be omitted and introduce components into the EIH at submultiples of the component frequency. B) a signal with a rapidly modulated envelope may contain turning points between zero crossings, introducing high-frequency artefacts into the EIH. (Note that the zero crossings are unaffected by either type of variation in the envelope.)

of the kth ZCPA bin is

$$\mathrm{ZCPA}[k] = \sum_{\iota=1}^{n} \log P_\iota .$$

(We neglect the $+1$ in the log argument for simplicity.) If the signal is scaled wholesale by a factor A, then the contribution to the histogram bin becomes

$$\mathrm{ZCPA}[k] = \sum_{\iota=1}^{n} \log AP_\iota = \sum_{\iota=1}^{n} \log P_\iota + n \log A.$$

Evidently, if $A \gg P$, then the chief influence upon the ZCPA spectrum becomes n: a simple count of the intervals, incorporating no amplitude information. We can now see that, for sufficiently intense input, the saturated ZCPA and EIH are identical, up to a scale factor. Despite exhibiting the same limiting behaviour, the ZCPA nevertheless remains superior to the EIH in a normal operating range: the ZCPA saturates less rapidly than the EIH (see Figure 2.20); and the ZCPA changes smoothly in response to intensity changes, whereas the EIH does not. It is important to note that it is not the ZCPA that brings about the saturation *per se*; rather, it is the log-compression of each peak prior to forming the histogram. Accordingly, if the ZCPA were computed directly from the peaks[1] (and the log-compression were perhaps performed on the ZCPA as a final step), then no saturation would result whatsoever.

The second set of problems which arise in the EIH but not the ZCPA go under the heading "spurious intervals" and result directly from interaction between the level crossing detectors and the signal envelope. The first type of spurious interval we may designate a *subharmonic interval*. These occur when modulation in the envelope of a harmonic causes some level crossings to drop out, introducing subharmonics

[1] this is equivalent to using $G(x) = cx$, where c is a positive constant.

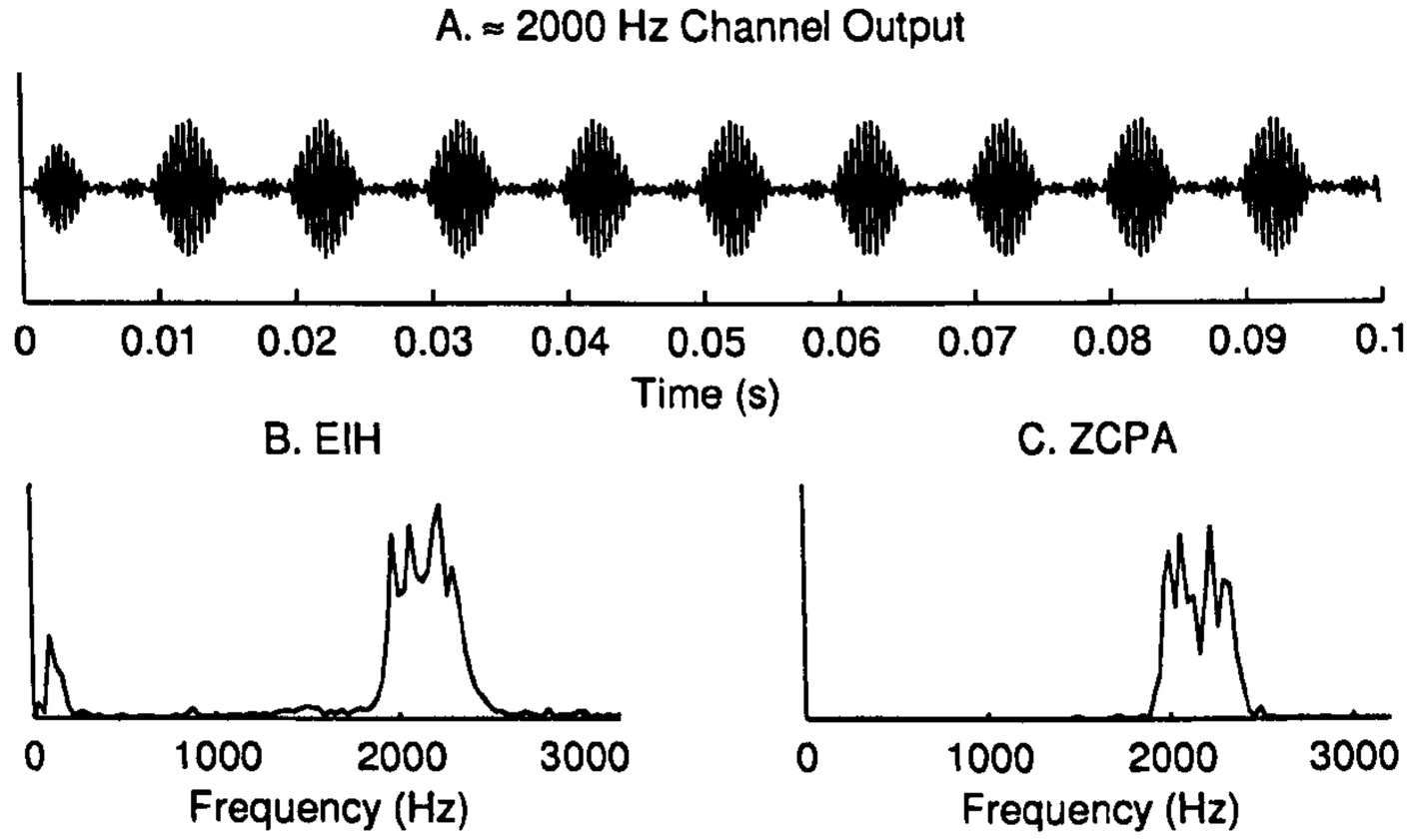


Figure 2.22: A) the output of a broad channel centred at 2 kHz in response to a harmonic complex with a 100 Hz fundamental frequency. The fine structure and envelope are 'phase-locked' to approximately 2 kHz and 100 Hz, respectively. B) a mixture of high-frequency harmonics contributes both short intervals (from the zero crossings) and period intervals (from the envelope) to the EIH. C) period intervals are absent from the ZCPA, which uses only zero crossings.

into the EIH (Figure 2.21A). A second type of spurious interval, which we designate *turn intervals*, emerges as the result of turning points between zero crossings, which split intervals equal to the harmonic period into shorter intervals (Figure 2.21B). In summary, subharmonic intervals are due to the omission of level crossings and produce low-frequency artefacts; conversely, turn intervals are due to the insertion of level crossings and produce high-frequency artefacts.

The third type of spurious interval we label *period intervals*. When a harmonic complex, such as the vowel sounds discussed above, forms the input to an EIH, several harmonics of the fundamental frequency are typically resolved in each filter, especially the broader filters covering the high-frequency region. The interaction of many harmonics in a channel gives rise to a periodic component in the envelope corresponding to the fundamental frequency (see, e.g., Figure 2.22A); and this has the potential to be captured by high-level crossing detectors as period intervals. This effect is observed in Figure 2.22B, which plots the EIH for a signal comprising four harmonics of a 100 Hz fundamental, namely, 2000 Hz, 2100 Hz, 2200 Hz and 2300 Hz. In addition to the high-frequency intervals and energy contributed by the harmonics themselves, the EIH registers a peak at 100 Hz, which is derived from the envelope of high-frequency filter outputs. This could be regarded as a low-level mechanism for demodulation of the envelope and may have some functional significance in the auditory system (Khanna and Teich, 1989).

The three types of spurious interval we have identified above are all generated by the level crossing detectors and thus are absent from the ZCPA, which extracts only zero crossing intervals. It is clear, for example, in Figure 2.21A(i) and B(i), that

whilst level crossing intervals are sensitive to local fluctuations in the narrowband envelope, the zero crossing intervals accurately convey the dominant frequency[1]. For the same reason, the ZCPA representation of four closely-spaced harmonics does not contain pitch intervals, as the EIH does; rather, the harmonics are displayed as a mass of high-frequency activity (Figure 2.22C). Consequently, the remark made above that "[weighting by] the peak amplitude can be likened to the joint contribution of a 'continuum' of level crossings detectors" applies only to *amplitude* encoding; a (theoretical) continuum of level crossing detectors and an upward zero crossing detector encode *frequency* in a markedly different manner.

Finally, *perturbation noise*—listed above as item (iii)—is another problem associated with the extraction of frequency information from level crossing intervals, to which Kim et al. (1999) draw attention. Their analysis assumes that level crossings have neither been deleted nor inserted (in other words, the spurious interval errors referred to above are not an issue); rather, the study considers the effect of additive Gaussian noise on level crossing times that, prior to the addition of noise, provide a useful frequency estimate. Under a high-SNR assumption, it is shown that adding noise samples displaces the level crossings from their initial times by a small amount. Specifically, if one assumes that the signal is a sinusoid with amplitude A and radial frequency ω added to a zero-mean Gaussian noise signal with variance σ_n^2, then the variance in the crossings of level l is given by

$$\sigma_{\mathsf{E}}^2(l) = \frac{\sigma_n^2}{A^2\omega^2(1 - l/A)^2}, \tag{2.10}$$

(Kim et al., 1999) with zero crossings providing a trivial case: $\sigma_{\mathsf{E}}^2(0) = \sigma_n^2(A\omega)^{-2}$. (The result in (2.10) is obtained by replacing the signal surrounding the crossing time with a first-order Taylor series expansion and considering the effect of additive noise on the intersection of the line with the time axis; see Kim et al. (1999) or Section 5.2.) Because the variance of level crossing perturbations has a global minimum when $l = 0$, we may safely conclude with Kim et al. that only *zero crossings* of the signal should be employed in frequency analysis, which is true of the ZCPA but not the EIH.

III. *Spatial-temporal Processing*

The third approach discussed here, *spatial-temporal processing*, detects the presence of signal components using time differences between neighbouring channel outputs. Algorithms in this category are relevant to the present study due, first of all, to the explicit reliance on temporal features of the signal, and secondly, to the auditory motivation claimed in their favour. This section briefly describes two such models: the *lateral inhibition network* (Shamma, 1985b) and the *fine structure spectrogram* (Dajani et al., 2005).

Biological lateral inhibition networks (LINs) are assemblies of nerves cells that act to sharpen the discontinuities in the spatial excitation patterns by a process of mutual suppression, and are present in most sensory systems. An auditory model presented

[1] see Kedem (1986).

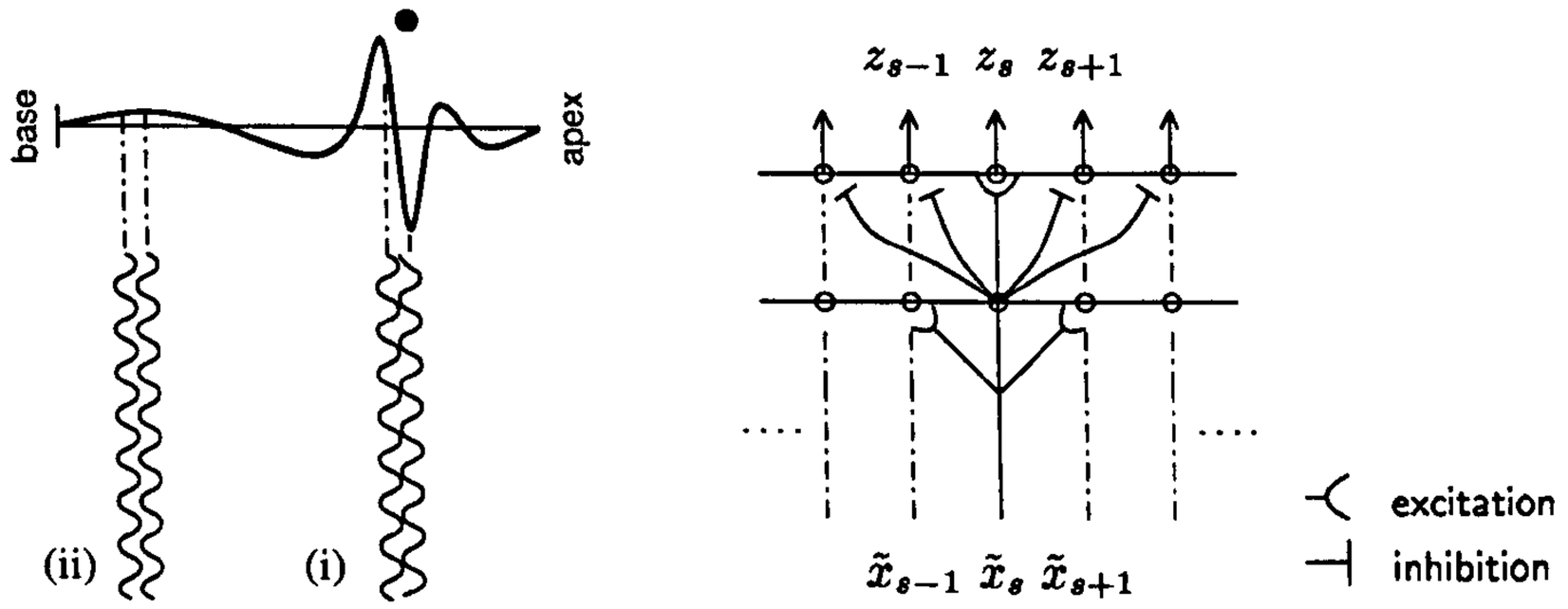


Figure 2.23: Processing in a lateral inhibition network. A) the phase of vibration changes more rapidly along the basilar membrane near resonant peaks [adapted from Shamma (1985a)]; B) the (non-recurrent) LIN exploits phase changes along the basilar membrane by implementing a form of spatial high-pass filter using simple, biologically-feasible units and inhibitory connections between the first and second layers to suppress the activity of in-phase regions [adapted from Shamma (1985b)].

by Shamma describes three stages of processing: analysis, transduction and reduction. The first two stages are familiar: "analysis" refers to cochlear filtering; "transduction" refers to the compression (possibly including half-wave rectification) and smoothing associated with the inner hair cells. In the original study (Shamma, 1985a,b), the initial stages were by-passed, and measurements were recorded directly from auditory fibres in the cat; only the reduction stage was simulated. In later studies (e.g., Wang and Shamma, 1994), all three stages were modelled. The details of the analysis and transduction steps are unimportant in this context; the only requirement is that the filterbank reproduce certain characteristics of the travelling wave along the basilar membrane.

The reduction stage uses spatial phase changes along the basilar membrane to detect spectral prominences. Small phase differences accumulate in the travelling wave away from the resonant point; towards the peak displacement, the travelling wave slows down and its spatial phase changes rapidly (Pickles, 1988). Figure 2.23A illustrates how two adjacent points respond (i) out-of-phase around the peak displacement, and (ii) in-phase in remote regions. Figure 2.23B shows the schematic for a short section of non-recurrent lateral inhibition network. The time-varying inputs to the LIN, $\{\tilde{x}_s\}$, are the outputs of the transduction stage—a process assumed to modify the basilar membrane motion in a way that preserves relative phase differences along the spatial axis, e.g., a static nonlinearity. The LIN processing itself is performed by two layers of units. Each LIN input excites a range of leaky integrator units in the first layer. The activation of a lower-level unit then *excites* the corresponding (non-leaky) unit in the layer above but *inhibits* its neighbours. (This is a non-recurrent LIN. A recurrent LIN consists of a single layer of units onto which inhibitory weights feed back.) We may now consider the effect of the inhibitory profile at sites removed from the resonant peak.

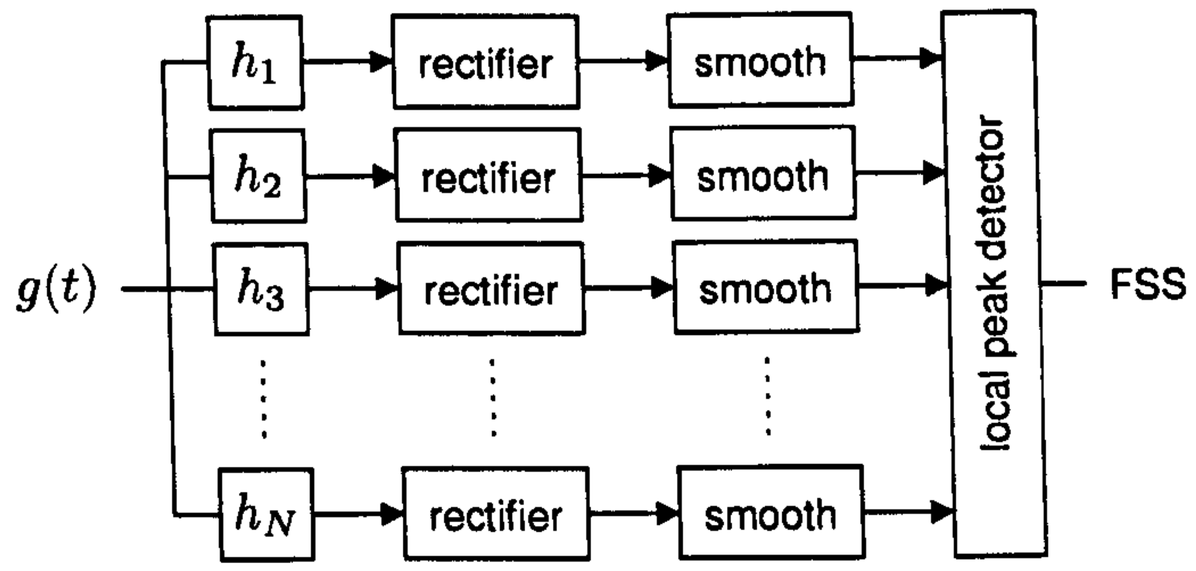


Figure 2.24: Fine structure spectrogram flow diagram.

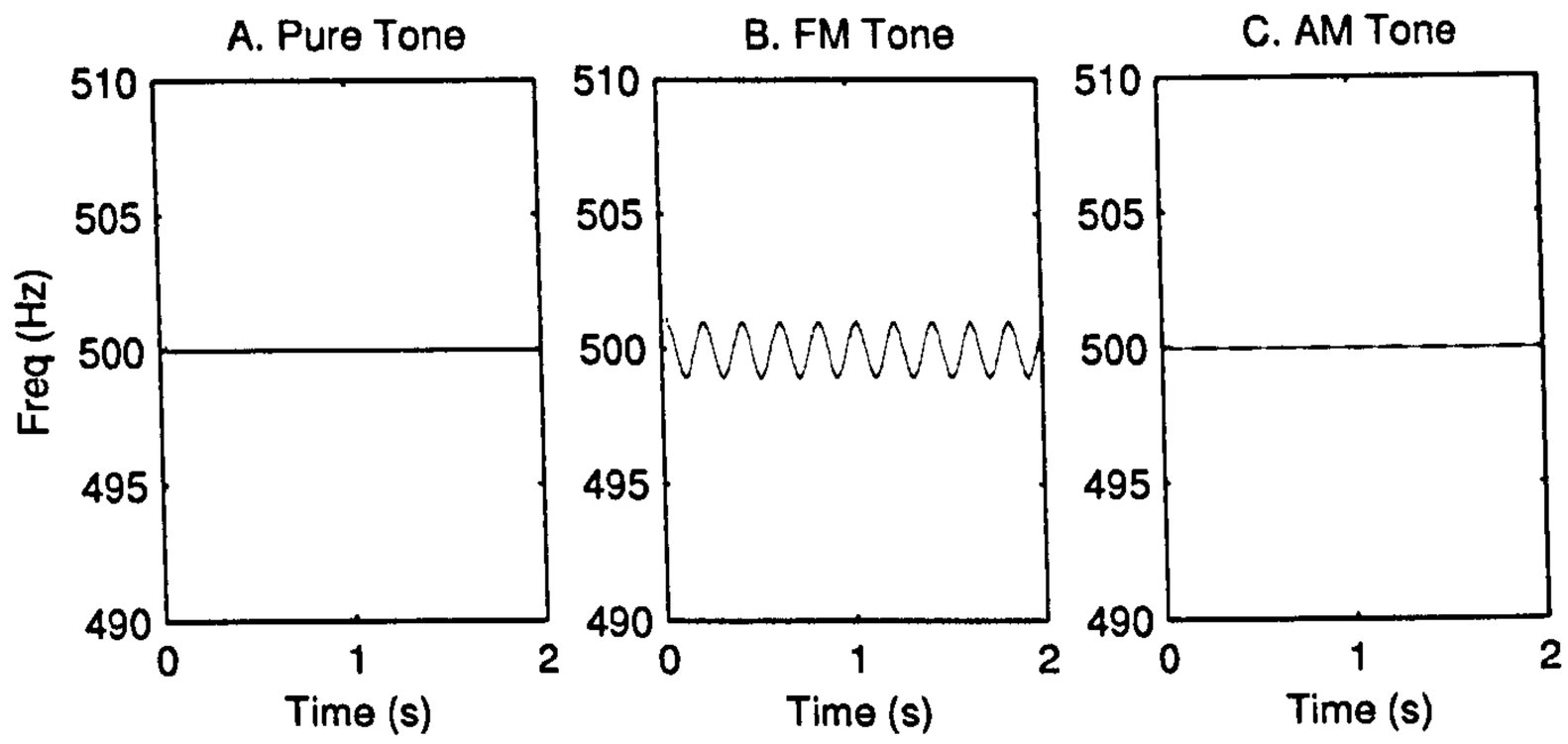


Figure 2.25: Fine structure spectrograms for A) a pure tone; B) a frequency-modulated tone; and C) an amplitude-modulated tone. *Cf.* Dajani et al. (2005), Fig. 2.

In these regions, closely-spaced points vibrate in phase, and inhibition acts to cancel the coincident patterns of neural activity. Turning to the converse case: the inhibitory profiles boosts—or simply fails to inhibit, depending on the implementation—activity surrounding resonances. In this way, LIN processing accentuates spectral peaks using solely timing information.

The second auditory-inspired approach to spatial-temporal processing discussed here is the fine structure spectrogram (FSS) proposed by Dajani et al. (2005). The fine structure spectrogram is not a temporal representation *per se* and hence demands only a cursory inspection. The stages in the production of the FSS are outlined Figure 2.24. The first block of processing is a densely-spaced bank of broadly overlapping filters. The output of each filter is full-wave rectified and smoothed, resulting in an approximation of the instantaneous subband envelope. These filters are linear-phase with equal group delay (including zero-phase as a special case) and peak gain. The final stage detects local peaks along the spatial axis (implicitly defined by the channel number) and plots them in a time-frequency space, taking the (x, y) coordinates from the sample time and channel centre frequency, respectively, and the colour value from the envelope.

Fine structure spectrograms for three types of elementary signal—a pure tone, a frequency-modulated (FM) tone and an (AM) amplitude-modulated tone—are shown in Figure 2.25. A pure tone at 500 Hz generates a peak in the wideband filter with the centre frequency closest to the tone and thus manifests itself in the FSS as a sharp, horizontal line. The spatial peak for an FM tone moves in a pattern resembling the modulating signal itself, provided that the frequency and depth of modulation are very much smaller than the filter analysis bandwidths. Finally, a carrier with an AM frequency considerably smaller than the analysis bandwidth is resolved in every filter as an AM signal, but the spatial peak consistently occurs at the carrier frequency. Such a signal is therefore represented in the FSS as a line of varying intensity in the colour map. Dajani et al. (2005) further demonstrate that the FSS provides a sharply-defined and meaningful depiction of AM-FM tones and full speech signals.

We may question whether the FSS can rightly be labelled a temporal representation, as it relies on a quantity allied to time-varying instantaneous power, rather than phase. Against this it may be argued that the even group-delay across the filters, coupled with the very frequent (e.g., sample-level) probing interval of the wideband filters, is itself a form of temporal processing. A key difference between the spatial processing of the FSS and the LIN, to which Dajani et al. also draw attention, is the manner in which the phase differences between filters are utilised: the FSS requires the absence of spatial phase delays along the filter bank in order to correctly demodulate signals; in contrast, the LIN exploits spatial phase delays in order to detect components and does not function without them.

2.2.5 Temporal Analysis in Computational ASA

The computer models presented above seek to transform an acoustic signal into the kind of neural representation that the brain might exploit in hearing, according to various contemporary theories of encoding. To model how the brain then proceeds to organise a mixture of sound sources into streams—to adopt the language of Bregman's auditory scene analysis framework (see Section 2.1.5 above)—requires a qualitatively new kind of computational approach. The field of study concerned with embodying the segmentation and grouping principles of ASA in a computer program is called *computational auditory scene analysis* (CASA). The last two temporal representations to be examined in this chapter—synchrony strands and the correlogram—are associated with CASA models; but before attending to these, a brief summary of some milestone developments in CASA is in order. For extensive reviews, see Wang and Brown (2006) and Cooke and Ellis (2001).

An early precursor to a CASA system was the speech separation algorithm developed by Parsons (1976) aimed at neutralising crosstalk on radio communications channels. This system operates on a frame-by-frame basis: harmonic peaks are detected in the Fourier frequency domain and compiled into a *peak table*; and, from here, fundamental frequencies are extracted using a histogram approach (Schroeder, 1968). Once the pitches of the talkers are available, the constituent voices are resynthesised from the peak tables using the pitch estimates as a guide. Although Parson's system is not based on auditory processing, it incorporates some important aspects of CASA:

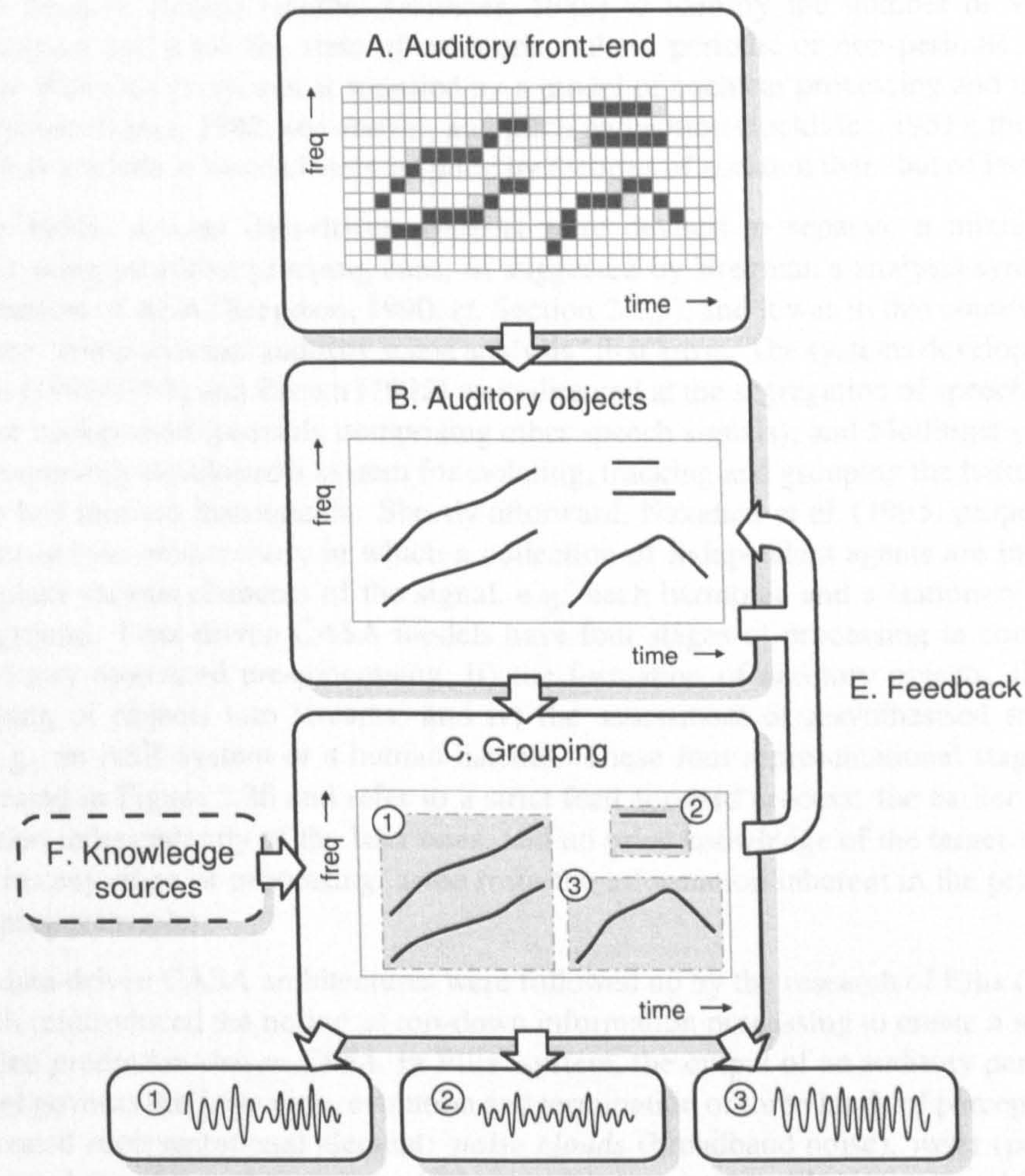


Figure 2.26: Flow chart illustrating the main steps in signal separation using CASA. A) a model of the auditory periphery typicially represents the signal as a two-dimensional array of pixels (e.g., FFT, ZCPA, EIH, LIN, GSD, FSS, wavelet) akin to a bitmap; B) the signal is decomposed into parametric auditory objects that are 'plotted' into an empty space; C) auditory objects are grouped together using Bregman's principles, such as common onset (as in 1,2) and frequency variation (as in 2); D) a resynthesis process allows the constituent signals to be recovered from the grouped components; E) residue and prediction-driven architectures incorporate a feedback pathway wherein the search for objects is informed by the current internal state of the model; and F) blackboard architectures cater for the inclusion of addition knowledge sources, such as a memory store.

acoustic elements are identified (harmonics) and then fused according to a grouping cue (common fundamental). Later, Weintraub (1985) developed a system that employed hidden Markov models (HMM) (Rabiner, 1989) to identify the number of voices in a mixture and track the state of each one: silent, periodic or non-periodic. The input to Weintraub's system is supplied by a model of cochlear processing and neural transduction (Lyon, 1982, see above), and pitch perception (Licklider, 1951); thus the system as a whole is more closely inspired by theories of audition than that of Parsons.

In the 1990s, several data-driven systems were devised to separate a mixture of sounds using primitive grouping cues, as suggested by Bregman's analysis-synthesis presentation of ASA (Bregman, 1990, *cf.* Section 2.1.5), and it was in this context that the term "computational auditory scene analysis" first arose. The systems developed by Cooke (1991/1993) and Brown (1992) were directed at the segregation of speech from a noise background (possibly comprising other speech signals); and Mellinger (1991) contemporarily developed a system for isolating, tracking and grouping the harmonics of pitched musical instruments. Shortly afterward, Nakatani et al. (1995) proposed a *residue-driven architecture*, in which a collection of independent agents are invoked to explain various elements of the signal, e.g., each harmonic and a stationary noise background. Data-driven CASA models have four stages of processing in common: i) auditory-motivated pre-processing, ii) the formation of auditory objects, iii) the grouping of objects into streams, and iv) the assessment of resynthesised streams by, e.g., an ASR system or a human listener. These four representational stages are illustrated in Figure 2.26 and refer to a strict feed-forward process: the earlier stages function independently of the later ones, and no prior knowledge of the target signals informs any stage of processing, aside from the information inherent in the primitive grouping principles.

The data-driven CASA architectures were followed up by the research of Ellis (1996), which reintroduced the notion of top-down information processing to create a scheme entitled *prediction-driven CASA*. In Ellis' system, the output of an auditory periphery model governs the formation, evolution and termination of three kinds of perceptually-motivated representational element: *noise clouds* (broadband noise), *wefts* (periodic wideband signals), and *transients*. This aspect is consistent with conventional CASA; the novel contribution lies in a *prediction-reconciliation engine*. Each element predicts the energy it will contribute to the next frame of input, according to various regularities: for example, a noise cloud is expected to add the same energy spectrum as it did to the preceding frame, and transients are expected to decay exponentially. Small deviations between the predicted and observed frames are used to modify the parameters of each element, which, in effect, hypothesises a changing element and implements a form of tracking. Failing this, the absence of expected energy generates a hypothesis that an element has disappeared, whilst the presence of unexpected energy signals the arrival of a new source that must be accounted for. It is the task of the reconciliation engine to choose amongst these hypotheses and to trigger an appropriate action, i.e., to create, update or discontinue representational elements. The prediction-reconciliation engine is implemented as a *blackboard architecture*, in which many independent knowledge sources, both data-driven and high-level (e.g., a database), execute actions and influence a globally-accessible "blackboard" of competing hypotheses. Wang and

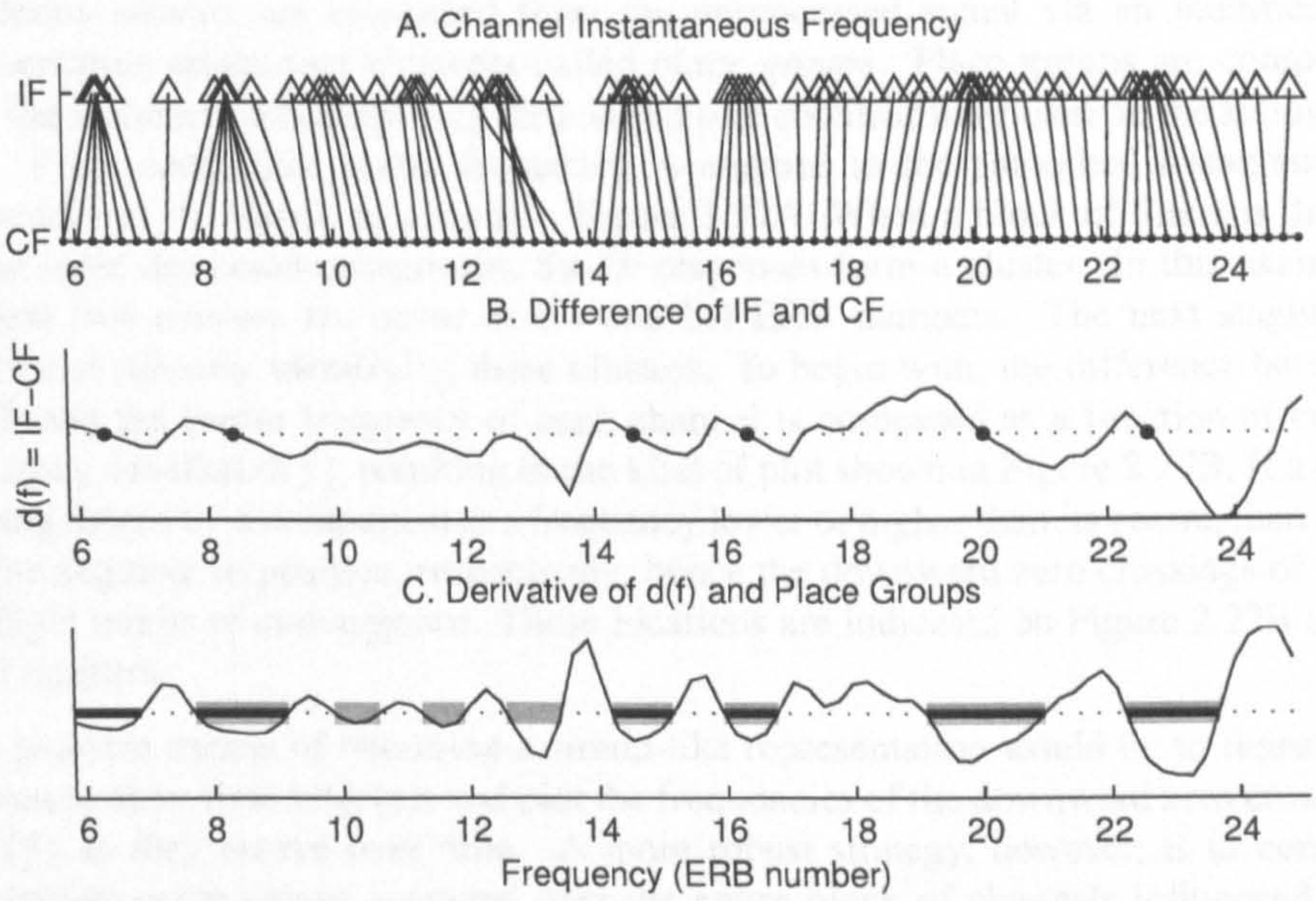


Figure 2.27: The main stages by which place groups are formed for the [oo] sound in the phrase "set white with P two soon". A) filter centre frequencies—shown here on an ERB scale—are mapped to smoothed instantaneous frequency estimates. The appearance of "bunches" signals the existence of dominant spectral components coinciding with formants. B) plotting the difference between the IF output of a channel and its centre frequency reveals the direction towards which its output gravitates—positive denotes upward in frequency, so a downward zero crossing accompanies each bunch in (A). C) the extent of each place group is determined by differentiating (B) and examining the width between the zero crossings, shown here as grey blocks. Black lines identify the groups that contain a zero crossing.

Brown (1999) have attempted to provide a physiological basis for the grouping and separation process using a circuit of model neural oscillators.

IV. *Synchrony Strands*

Cooke's CASA model computes a collection of auditory objects from the raw signal called *synchrony strands*, each of which represents the "time-frequency behaviour of a single spectral component (e.g. harmonic or speech formant)" (Cooke, 1991/1993). Each strand effectively traces one part of the signal deemed to have arisen from the same source, according to the principle of good continuity, described in Section 2.1.5. A complete set of strands constitutes a description of the entire signal. The computation of synchrony strands draws together aspects from all three temporal representations discussed in Section 2.2.4: joint-synchrony, in-channel (i.e., zero crossing) and spatial-temporal processing. We shall be in a better position to justify these comparisons after a closer examination synchrony strands themselves.

Synchrony strands are computed from the unprocessed signal via an intermediate representation containing elements called *place groups*. Place groups are computed from the instantaneous frequency (IF) output of a cochlear filter bank in the following way. First, each filter centre frequency is mapped to the smoothed instantaneous frequency[1] of its output, as shown in Figure 2.27A. When a block of filters is driven by the same dominant component, the IF responses form a cluster. In this example, the first two clusters are occur at 6.4 and 8.4 ERB numbers. The next stages are directed at robustly identifying these clusters. To begin with, the difference between the IF and the centre frequency of each channel is computed as a function of centre frequency, labelled $d(f)$, resulting in the kind of plot shown in Figure 2.27B. If a filter is being driven by a component at a frequency lower or higher than its centre, then $d(f)$ will be negative or positive, respectively; hence the downward zero crossings of $d(f)$ highlight points of convergence. These locations are indicated on Figure 2.27B using solid markers.

One possible means of obtaining a strand-like representation would be to repeat this process at short time intervals and plot the frequencies of the downward zero crossings in $d(f)$ as they evolve over time. A more robust strategy, however, is to compute parameters using values averaged over the entire block of channels influenced by a dominance. Cooke (1991/1993) defines a place group as the range of frequencies between any consecutive local peak and trough in $d(f)$, after smoothing with a Gaussian kernel; or, equivalently, the range of frequencies enclosed by consecutive negative-going and positive-going zero crossings in the smoothed derivative of $d(t)$. (An optional further step might involve eliminating place groups that do not correspond to a downward zero crossing in $d(t)$. This action would remove some 'noisy' place groups, at the expense of deleting genuine place groups associated with weak components in the vicinity of much stronger spectral prominences.)

Two further stages are undertaken to convert place groups into synchrony strands. First, place groups are aggregated to form longer groups using frequency trajectories as a guide (Cooke, 1991/1993, page 42). Second, a set of attributes is calculated for each group, e.g., frequency, *dominance* (the breadth of the place group), amplitude and AM. Once all the strands have been computed, the grouping stage can commence. However, we can leave aside these later stages, as it is the initial formation of place groups that is most relevant to temporal processing. It is interesting to note, in closing, that synchrony strands incorporate aspects from all three types of timing representation presented in the previous section. Synchrony strand processing employs a similar idea to Seneff's joint-synchrony representation, as it measures the difference between the IF output of a channel and its centre frequency. It also resembles the EIH and ZCPA in two regards: i) the zero crossing intervals of a narrowband signal and its phase derivative convey the same quantity, namely, instantaneous frequency; and ii) both methods employ synchrony capture as a mode of detection[2]. Finally, we may firmly include the synchrony strand technique amongst other spatial-temporal processing methods, e.g., the LIN and FSS, as it evidently relies upon the concurrent output of many filters.

[1] For details concerning the calculation and smoothing of the IF, see Cooke (1991/1993, page 40).
[2] See the related work on *in-synchrony bands* by Ghitza (1988).

V. *Autocorrelation*

Several CASA models employ an autocorrelation technique to detect synchrony in an auditory filter as an alternative to instantaneous frequency (Brown and Cooke, 1994; Brown, 1992). These models make extensive use of the *correlogram*: a two-dimensional representation whose rows contain the autocorrelation function measured in each channel at given time instant (Slaney and Lyon, 1990). It has the discrete-time definition [adapted from Wang and Brown (2006)]:

$$\underset{\tilde{x}}{\text{acf}}[n, s, k] = \frac{\sum_{\iota=0}^{\infty} w[\iota]\tilde{x}_s[n-\iota]\tilde{x}_s[n-k-\iota]}{\sum_{\iota=0}^{\infty} w[\iota]\tilde{x}_s^2[n-\iota]}, \tag{2.11}$$

in which $\tilde{x}_s[n]$ is the output of the analysis-transduction model at sample time n, in the channel indexed s; k is a sampled time lag; and $w[\cdot]$ is a tapered window function, included to localise the representation in time. Note that the correlogram is normalised to unity at lag time $k = 0$; however, the formulation in (2.11) does *not* guarantee that all values in the correlogram are less than one. A column-wise average of the correlogram produces a *summary correlogram*, which highlights every vertical ridge as a peak.

A frame-based approach to computing the autocorrelation extracts a frame of samples from channel s concluding on sample n into a vector $\tilde{\mathbf{x}}_{n,s}$, and performs the following sequence of operations:

$$\underset{\mathbf{x}}{\text{circ-acf}}[n, s, k] = \frac{\text{IDFT}_k\{|\,\text{DFT}\{\tilde{\mathbf{x}}_{n,s}\}|^2\}}{\tilde{\mathbf{x}}_{n,s} \cdot \tilde{\mathbf{x}}_{n,s}}. \tag{2.12}$$

This procedure is based on a circular convolution and hence is equivalent to using (2.11) and assuming that the signal consists of the same frame concatenated endlessly. To avoid discontinuities at the frame boundaries, it is advisable to apply a smooth, tapered window to the frame prior to the DFT. Alternatively, an autocorrelation function can be obtained from the frame directly by convolving the time-domain signal with a reversed version of itself and dividing by the frame energy. This approach is equivalent to (2.11), if the signal is assumed to be zero everywhere except within the frame boundary. Both frame-based techniques guarantee a maximum at $k = 0$.

Figures 2.29A–C show, from top to bottom: a plot of the simulated auditory nerve activity—often referred to as a *ratemap*, a correlogram and a summary correlogram for a synthetic vowel sound `[ar]`, with a fundamental frequency of 106 Hz. The magnitude spectrum of the vowel is plotted separately in Figure 2.28A. The ratemap is generated by applying a half-wave rectification and compression[1] to the output of a 32-channel gammatone filter bank, i.e.,

$$\tilde{x}_s(t) = \begin{cases} \sqrt{x_s(t)} & x_s(t) > 0 \\ 0 & \text{otherwise} \end{cases}, \; 1 \leq s \leq 32. \tag{2.13}$$

The correlogram in (B) is computed from 50 ms Hamming-windowed sections of $\tilde{x}$, starting at 100 ms, using the DFT method (2.12). A peak is present in every channel at

[1] A cube-root or logarithmic compression function is often used to produce ratemaps.

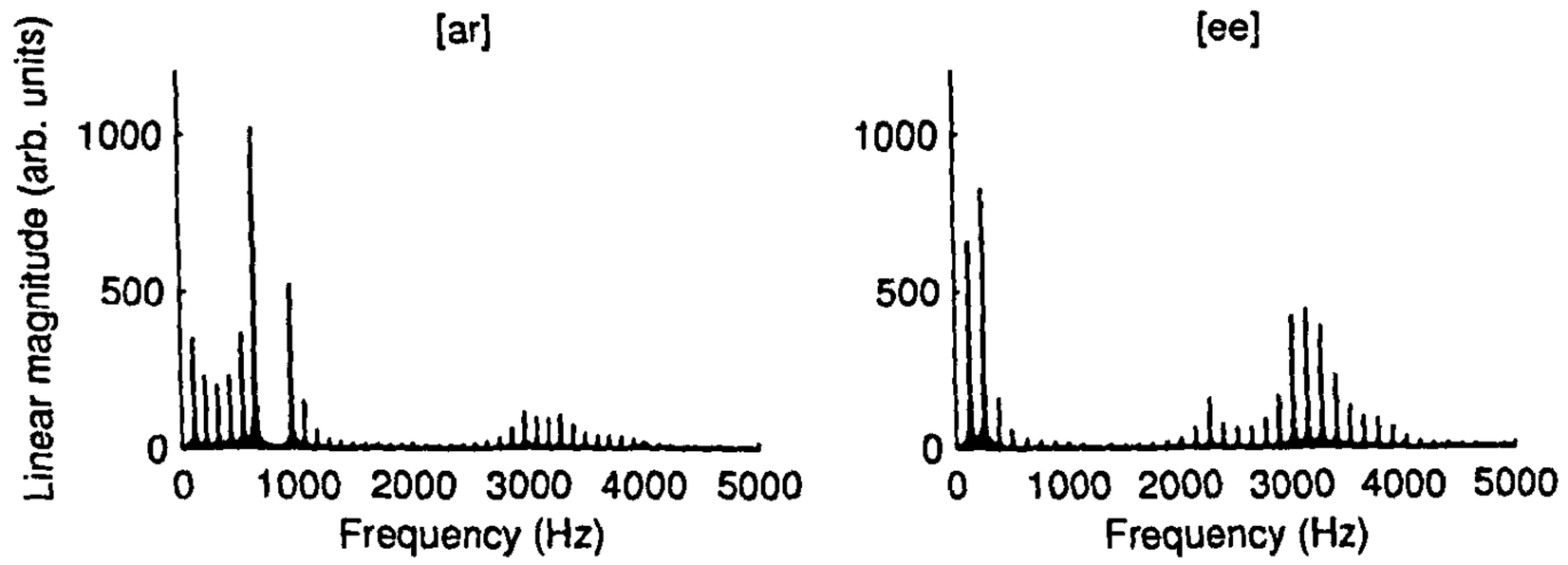


Figure 2.28: Magnitude spectrum for two artificial vowel sounds, [ar] and [ee], synthesised for, and used in a study conducted by, Summerfield and Assmann (1991).

approximately 9.4 ms, and the alignment of peaks creates the impression of a vertical ridge in the correlogram. In Figure 2.29C, the 9.4 ms peak is indicative of a 106 Hz fundamental frequency, and a second, weaker peak at 18.8 ms coincides with a time lag equal to two fundamental periods. (If the analysis window is lengthened, and the signal has a stationary pitch, peaks continue to recur at multiples of the fundamental period.)

The right-hand column of plots in Figure 2.29, D–F, display the ratemap, correlogram and summary correlogram after the vowel sound [ar] has been mixed with a second vowel sound, [ee], with a fundamental frequency of 126 Hz. (The signals have equal power, and the magnitude spectrum for the vowel [ee] is plotted in Figure 2.28B.) The correlogram (2.29E) now contains a few vertical ridges at various time lags, each localised to a different block of filters. From inspection, the channels with centre frequencies greater than 2200 Hz show broad ridges at 7.9 ms, 15.8 ms and 23.7 Hz—evidence for a 126 Hz pitch; channels in the range 500–2000 Hz continue to exhibit distinct ridges at 9.4 ms and 18.8 ms—evidence of a 106 Hz pitch. A visual assessment of the channels with frequencies below 500 Hz provides no evidence of a strong pitch component; however, there is clearly no longer any alignment at 9.4 ms, as there was in Figure 2.29B. The origin of these synchronised blocks can be traced to the magnitude spectra of each vowel, shown in Figure 2.28. In a given frequency band, the correlogram registers the pitch of the vowel with the dominant formant(s) in that region. For instance, [ar] has a prominent formant at 1 kHz, which is absent in [ee], so the region surrounding 1 kHz is synchronised to 106 Hz in the correlogram.

Subsequent research has sought to automate the process of identifying synchronised regions in the correlogram. One such approach is the *correlation map*: a mid-level representation in Brown's CASA model that forms periodicity groups by progressively merging channels into blocks according to a similarity metric (Brown, 1992). A simpler and conceptually equivalent approach computes the correlation coefficient for adjacent channels in the correlogram and reveals synchronised blocks by applying a threshold (Wang and Brown, 1999).

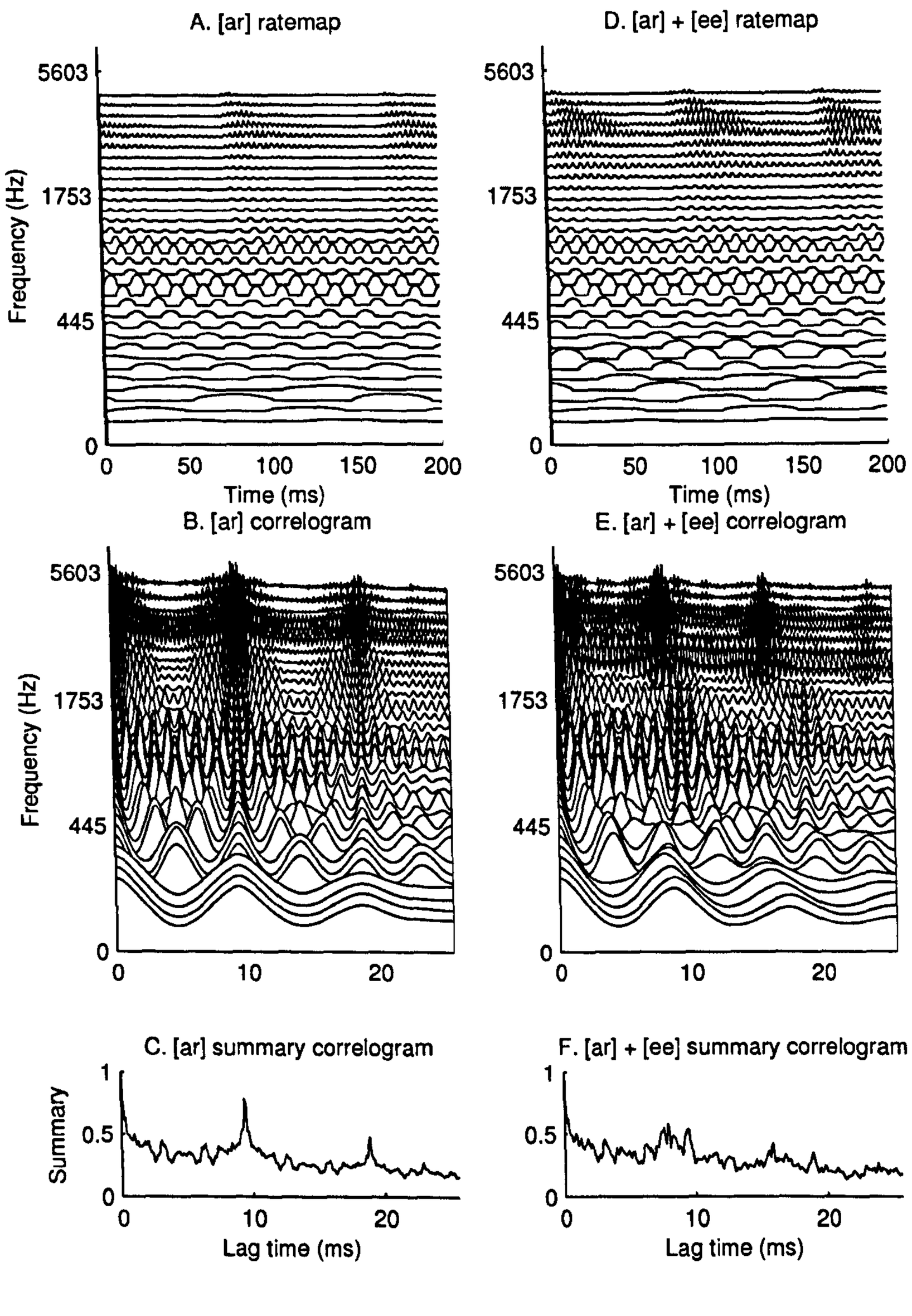


Figure 2.29: A) 32-channel rate map for the vowel sound [ar]; B) correlogram for the [ar] vowel; C) summary correlogram from (B). D–F) 32-channel rate map, correlogram and summary correlogram for the [ar]+[ee] vowel mixture. The magnitude spectra of the individual vowels are plotted in Figure 2.28.

2.3 Summary

For more than fifty years, experimenters have been able, in principle at least, to detect an acoustic signal by measuring the activity of a single auditory nerve fibre with a microelectrode. Physiological studies of the auditory-neural response to pure tones have convincingly demonstrated that a signal is encoded along a tonotopic axis by both the *average firing rate* (up to the point of saturation) and the *synchronised pattern of firing*. The average firing rate is related to the mechanical excitation of a place on the basilar membrane by energy in a critical band, whereas the temporal firing pattern reflects the fine structure of the stimulus, and captures both its frequency and AM/FM characteristics.

Conventional passive sonar receivers operate on a similar principle to rate coding. The resonant character of a place on the basilar membrane is comparable to a tuned analogue or digital band-pass filter. The firing rate of inner hair cells monotonically increases with peak displacement of the BM, and thus conveys a quantity similar to a non-linear compression of the filtered signal envelope. A simplified model of rate coding, in which the brain detects a tone by thresholding the firing rate, is qualitatively identical to passive narrowband sonar; only the orders of magnitude differ.

The remaining chapters carefully examine what form a passive narrowband sonar based on a simplified model of *temporal* coding might take. Just as a compressive function of the band-pass signal envelope has served the signal processing community as a simple model of peak displacement and compression on the BM, so the zero crossings of the band-pass signal have come to be identified with auditory nerve discharges in the auditory modelling literature. Furthermore, temporal theories that rely on the time interval between nerve discharges have translated into signal processing methods that measure zero crossing intervals.

Chapter

3

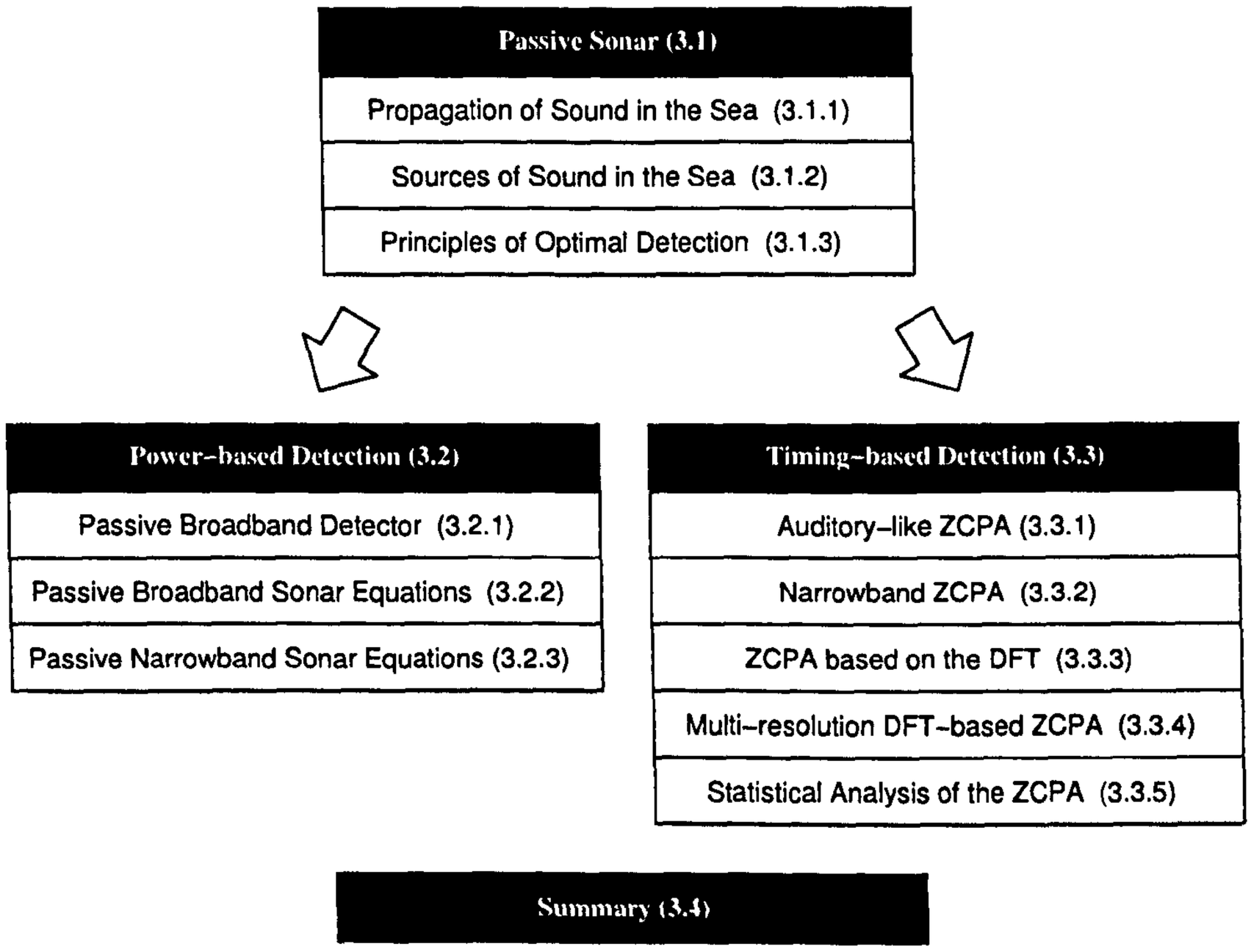
Auditory-motivated Sonar Displays

In the opening chapter we observed that passive sonar and the faculty of hearing employ broadly similar strategies, insofar as both infer information about the objects present in the environment according to some kind of time-frequency analysis of the sounds they produce. The preceding chapter described how the human ear analyses acoustic signals and concluded with a detailed review of five computational models of hearing, in particular, models inspired by temporal coding theories of audition. These models reflect a diverse range of opinion concerning how the timing of auditory nerve spikes encode information: spatial phase changes along the basilar membrane, inter-spike intervals, an average or time-varying measure of synchrony, and autocorrelation can all, in principle, be used to detect narrowband signals. In this chapter, we shall investigate the potential for auditory-inspired signal processing algorithms to assist narrowband sonar detection.

The first section in this chapter reviews those aspects of sonar which are common to power-based and timing-based processing. This account includes the propagation of sound in the sea (§3.1.1), a brief survey of common sound sources (§3.1.2), and a general overview of Bayesian hypothesis testing (§3.1.3). Rather than burden the text with a citation for every new term, the reader is directed to the chief sources of the material on acoustics, which include, in order of priority, Burdic (1984), Kinsler et al. (2000), Waite (1998) and Wright (2005), wherein a more complete and formal discussion can be found. Similarly, the optimal detection material draws heavily on Burdic (1984) and Whalen (1971).

The second and third sections describe two techniques for detecting the presence of a sound source mixed with a noise background received at a hydrophone array: one based on power (§3.2), the other on timing (§3.3). In the conventional, power-based approach, the recorded underwater sound is passed through a linear filter, and the power at the output of the system is used to decide whether a signal is present. The challenge,

Chapter 3 Outline



in many cases, is to choose the linear system that maximises detection performance according to some preferred criterion. In practice, the detection and classification decisions are made by human operators, who make use of spectrogram-like displays to visualise the power in multiple channels. However, the theoretical analysis of a sonar's performance normally refers to an automated procedure.

The third section examines the possibility of adapting the zero crossing with peak amplitudes (ZCPA) representation to suit sonar applications. This project faces two significant challenges. First, it seems clear from the outset that the relatively low frequency resolution (i.e., wide bandwidth) of an auditory filter bank will be difficult to reconcile with attempts to detect tonal components at very low signal-to-noise ratios. Ultimately, the solution should mimic the style of processing observed in the auditory pathway *and* satisfy the practical requirements of a narrowband sonar application. Second, careful thought must be given to the problem of comparing the performance of a timing-based detector with that of a traditional, power-based solution, especially if the two approaches process the signal in a radically dissimilar manner.

3.1 Passive Sonar

3.1.1 Propagation of Sound in the Sea

The propagation of a sound wave through the sea is determined by the physical laws that apply in a volume of sea water at any given instant. The discussion of wave motion in this section centres on two ideas conveyed in the preceding statement: the shape of a wave in space at a fixed time, and the changing properties of a wave in a fixed space. In seeking to understand more advanced concepts, we shall start with a very simple, and somewhat unrealistic, model of sound propagation and gradually refine it towards a more useful explanation of the way waves actually behave in the sea.

Plane Waves

Our discussion of waves opens with four basic working assumptions: i) that water extends infinitely in all directions; ii) that the medium is uniform in all aspects, with equal mean pressure (i.e., there is no gravity), and constant density and temperature; iii) the motion of particles in the medium is very small; and iv) no energy is lost from the system once introduced. With (i)–(iv) satisfied, we must now define some variables: a pressure scalar field p, a particle position vector field $\mathbf{x}$, and two constants ρ and K, which respectively denote the density and bulk modulus of elasticity of the medium. In a pressurised medium, every element exerts an outward force (equal to the product of pressure and area) on its neighbours. From Newton's Second Law, it follows that

$$\nabla p = -\rho \frac{\partial^2 \mathbf{x}}{\partial t^2} \equiv -\rho \frac{\partial \mathbf{u}}{\partial t}, \tag{3.1}$$

where $\mathbf{u}$ is particle velocity. This result (3.1) is *Euler's linear equation*, and it furnishes us with the first physical law that must apply at every point and time in the system. A second relationship is described by the *linear continuity equation* and says that the compression of an element and the outward force it exerts are directly proportional to each other, i.e.,

$$\frac{dp}{dt} = -K \nabla \cdot \mathbf{u}, \tag{3.2}$$

where K is the bulk modulus of elasticity, and we have differentiated both sides with respect to time. Combining the divergence of (3.1) and the time derivative of (3.2) and removing $\mathbf{u}$ results in the three-dimensional wave equation expressed in terms of pressure:

$$\frac{1}{c^2} \frac{d^2 p}{dt^2} - \nabla^2 p = 0, \tag{3.3}$$

where $c^2 = K/\rho$. Solutions to (3.3) take the form

$$p(t, \mathbf{r}) = p_1(t - \hat{\mathbf{k}} \cdot \mathbf{r}/c) + p_2(t + \hat{\mathbf{k}} \cdot \mathbf{r}/c) \tag{3.4}$$

where $\mathbf{r}$ is a position vector (x, y, z). Here, $p_1(\cdot)$ is an arbitrary function of one variable representing a plane wave moving normal to the unit vector $\hat{\mathbf{k}}$ with speed c m/s, and $p_2(\cdot)$ is another arbitrary function of one variable representing a plane wave moving in the opposite direction at the same speed.

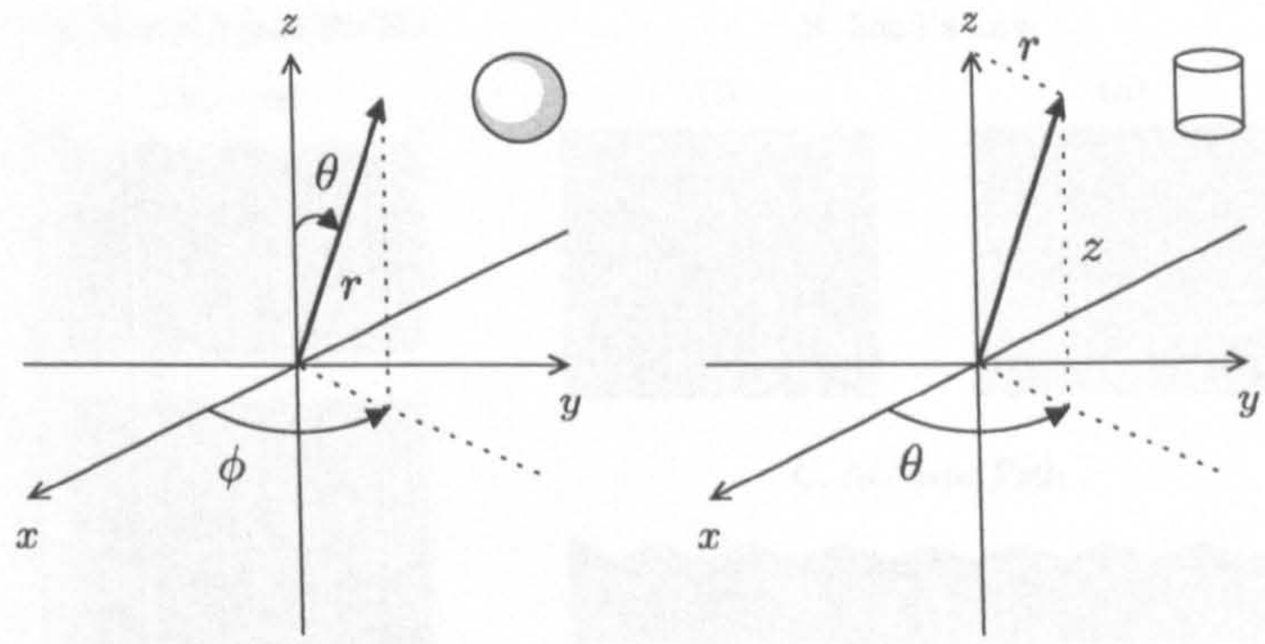


Figure 3.1: Two curvilinear coordinates systems: A) spherical coordinates (r, θ, ϕ) and B) cylindrical coordinates (r, θ, z) (Kinsler et al., 2000). Note the usage of the radial coordinate, r, in spherical and cylindrical coordinates: in the former, r measures the distance from the origin; in the latter, it measures the distance from the z-axis.

Spherical and Cylindrical Waves

In order to describe waves that propagate in a non-planar fashion, we replace the coordinate lines along which $\mathbf{x}$ measures (i.e., x, y and z) with new coordinate lines which map the space in a different way, e.g., in concentric circles (*cf.* Figure 3.1). The linear Euler and continuity equations still apply, only now they are enforced along curves. The linear wave equation expressed in a spherical coordinate system is identical to (2.3), except the Laplacian operator, ∇^2, is adjusted to operate along spherical, rather than Cartesian, coordinate lines. For a time-varying pressure at the origin, the linear wave equation can be shown to have the compact form

$$\frac{1}{c^2}\frac{d^2(rp)}{dt^2} - \frac{\partial^2(rp)}{\partial r^2} = 0, \tag{3.5}$$

where r is the distance from the origin, and admits physically-realisable solutions of the form

$$p(t, r) = \frac{p_1(t - r/c)}{r}, \tag{3.6}$$

in which $p_1(\cdot)$ is an arbitrary function representing a spherical pressure wave propagating away from the origin with speed c m/s. The key result here is that wave pressure must decay with distance from the source. The mean power per unit area, or *intensity*, of a spherical wavefront of radius r metres is given by

$$I_{sph}(r) = \frac{\Pi_{sph}}{4\pi r^2}, \text{ Watts/m}^2, \tag{3.7}$$

where Π_{sph} is the average power passing across a reference sphere of one metre radius. If the source at the origin is harmonic, with amplitude A_m at a one metre distance from the origin, then $\Pi_{sph} = 2\pi A_m^2/(\rho c)$.

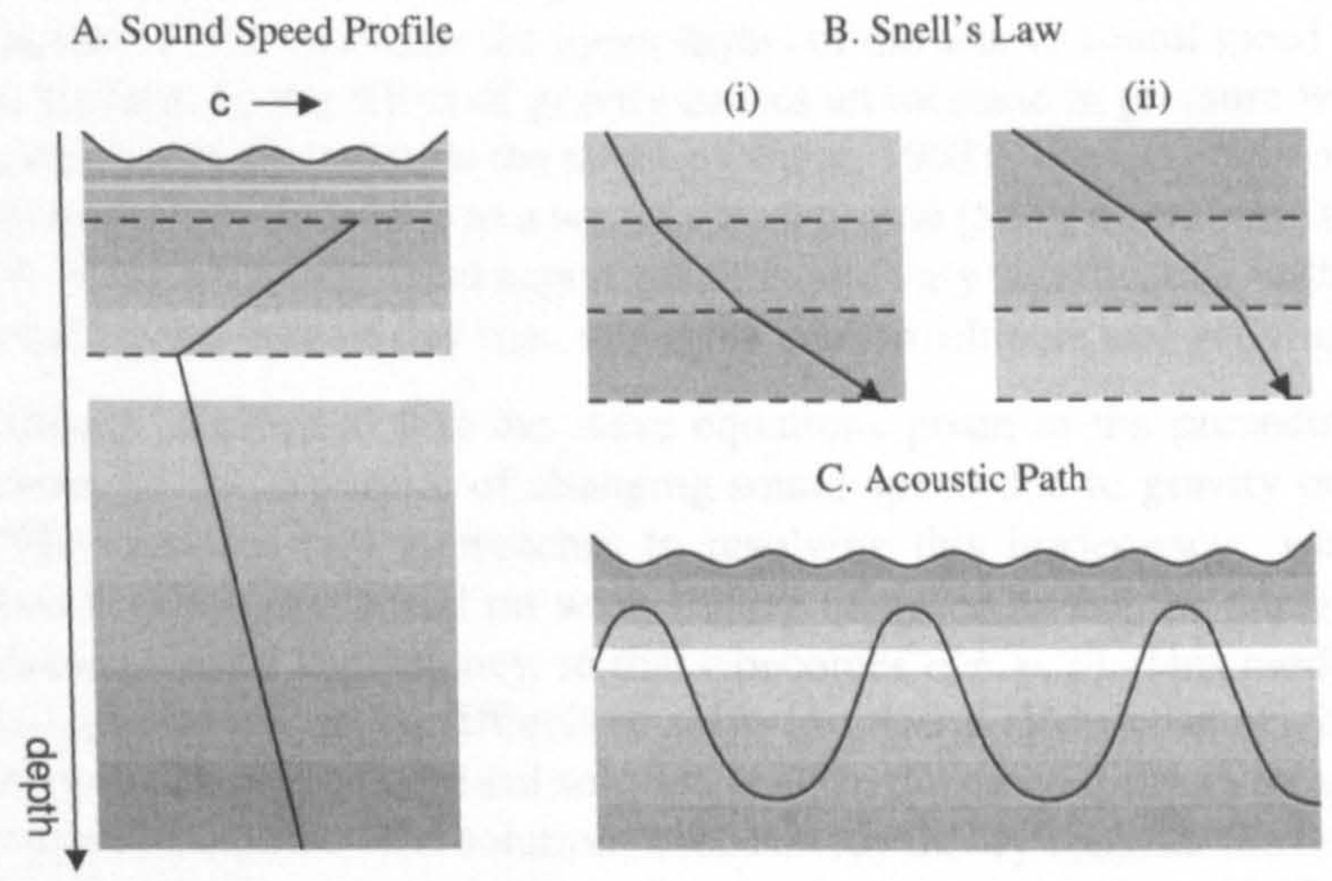


Figure 3.2: Sketches of various refraction effects. Darker shades correspond to higher sound speed. A) a typical sound speed profile; B) a (i) positive or (ii) negative SSP gradient causes a ray to refract towards or away from the boundary, respectively. C) the shape of a sound speed profile (A), combined with Snell's law (B), can trap sound in an acoustic path.

For cylindrical waves, the linear wave equation assumes the form

$$\frac{1}{c^2}\frac{d^2p}{dt^2} - \frac{\partial^2 p}{\partial r^2} - \frac{1}{r}\frac{\partial p}{\partial r} = 0, \tag{3.8}$$

where r now represents the distance along coordinate lines perpendicular to the z-axis, as shown in Figure 3.1B. Unlike plane and spherical waves, the analytic solution to (3.8) is non-trivial; intuitively, however, the solutions represent waves diverging from the z-axis (Kinsler et al., 2000). Once again, intensity decreases with distance:

$$I_{cyl}(r) = \frac{\Pi_{cyl}}{2\pi r h}, \text{ Watts/m}^2, \tag{3.9}$$

where Π_{cyl} is the average power passing across a unit cylinder of height h. The spherical and cylindrical intensity formulae will be revisited in connection with the sonar equations discussed in Section 3.2.2.

Refraction

The equations derived above govern the propagation of a wave under a set of idealised and somewhat unrealistic conditions, which we enumerated at the outset. It is arguably assumption (ii)—the absence of gravity and an isothermal medium—that causes the greatest departure between the model and reality, so we shall address this aspect first. The speed of sound, c, was taken to be constant throughout the seawater; in fact,

the speed of sound in the sea varies appreciably, particularly with depth. The cause of variation of sound speed with depth can be sketched in terms of two principal, opposing factors: 1) the sun heats the upper layers of the sea, so sound speed *increases* towards the surface; 2) the effect of gravity causes an increase in pressure with depth, so sound speed *decreases* towards the surface (Waite, 1998). The net effect of changes in temperature and pressure lead to a sound speed profile (SSP) resembling the plot in Figure 3.2A. SSPs are determined experimentally and vary significantly with a host of factors, including the time of day (i.e., sunlight), season, climate and geography.

We have already mentioned that the wave equations given in the preceding section fail to account for the influence of changing sound speed due to gravity or sunlight. Waite (1998) identifies two approaches to resolving this inadequacy: *wave theory* and *ray theory*. Solutions based on wave theory continue to use the wave equation, now introducing spatial dependency, so that c becomes $c(x, y, z)$. This modified wave equation is considerably more difficult to solve in general (Kinsler et al., 2000, page 135) but offers a complete, analytical solution at all frequencies and may be applied in a number of useful scenarios. The solutions based on ray theory consider the refraction of an acoustic ray at the boundaries between layers of different sound speed. Here, Snell's law (Burdic, 1984, page 114) is used to compute the angle of refraction at the interface of fluid layers. A point on a wavefront passing between two media is refracted, i.e., bent, towards the boundary if the sound speed increases, as in Figure 3.2B(i), and away from the boundary if the sound speed decreases, as in Figure 3.2B(ii). Where the SSP of the sea exhibits a steep negative gradient (due to the diminishing effect of the sun), followed by a gradual positive gradient (due to gravity) sound waves may become 'trapped' in a cycle of upward and downward refraction. This gives rise to an *acoustic channel*, along which sound can propagate undisturbed over many thousands of kilometres (Figure 3.2C) (Burdic, 1984).

Surfaces

The previous section augmented the model of sound propagation to account for the change in sound speed with depth, at least on a qualitative level. We now question assumption (i)—that the sea continues infinitely in all directions. Whilst many practical modelling scenarios allow that the sea possess infinite extent in the horizontal plane, clearly the same cannot be said for the vertical axis, as abrupt changes occur at the sea surface and the sea floor. In this section we shall outline some of the main effects.

The transition from water to air offers very little impedance against a pressure wave. A crude model of the sea surface is provided by a *pressure release surface*, along which the pressure is zero. If the surface is a horizontal plane at depth z_s, the linear wave equation is now subject to the boundary condition $p(t, x, y, z_s) = 0$. Solutions for a general plane wave then take the form

$$p\left(t + \frac{k_1 x}{c} + \frac{k_2 y}{c} + \frac{k_3(z - z_s)}{c}\right) - p\left(t + \frac{k_1 x}{c} + \frac{k_2 y}{c} + \frac{k_3(z_s - z)}{c}\right), \quad (3.10)$$

where $p(\cdot)$ is an arbitrary function of one variable. Figure 3.3A shows the interaction of a wave with a pressure release surface. Note that the reflected wave is a phase-inverted

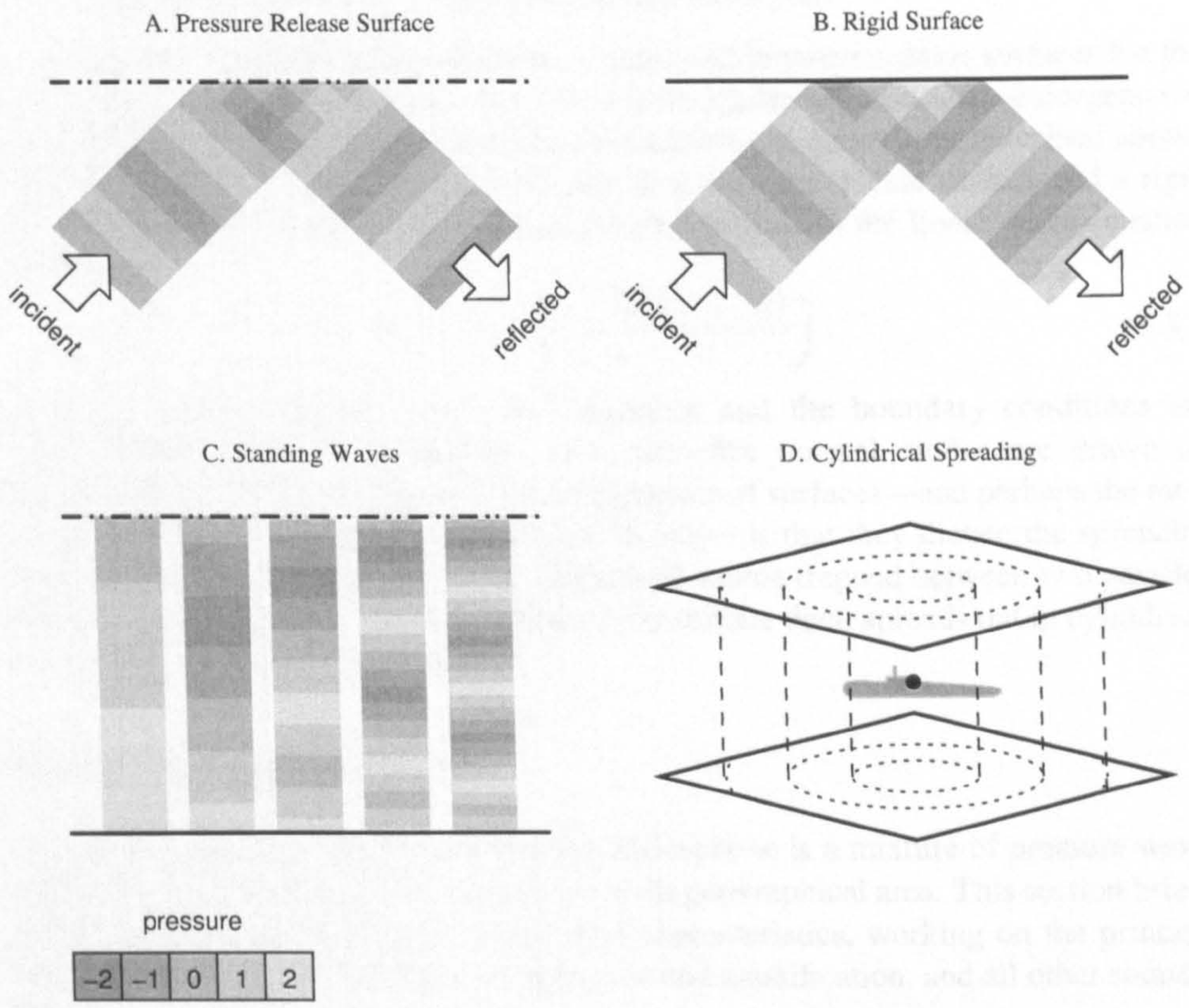


Figure 3.3: Sketches of various surface effects. Lighter shades correspond to higher pressures. A) reflection of a (partially drawn) plane wave from a pressure release surface; B) reflection of a plane wave from a rigid interface; C) the first five normal modes of standing waves between a soft and hard boundary; D) intuition underpinning the basis of cylindrical spreading in a layer.

mirror image of the incident wave, and that the addition of the two waves maintains zero pressure along the boundary at all times.

On the sea floor, the opposition to the flow of the water can be modelled by introducing a *rigid boundary*, that is, a surface at which the normal particle velocity component is zero. If the sea bed is a horizontal plane with depth z_d and normal vector $\hat{\mathbf{z}}$, then we enforce the boundary condition $\mathbf{u}(x, y, z_b) \cdot \hat{\mathbf{z}} = 0$. Solutions for a general plane wave take the form

$$p\left(t + \frac{k_1 x}{c} + \frac{k_2 y}{c} + \frac{k_3(z - z_b)}{c}\right) + p\left(t + \frac{k_1 x}{c} + \frac{k_2 y}{c} + \frac{k_3(z_b - z)}{c}\right), \quad (3.11)$$

where $p(\cdot)$ is an arbitrary function of one variable. Figure 3.3B shows the interaction of a wave with a rigid surface. The incident and reflected waves are now in-phase, so that the pressure at the boundary is doubled. However, the pressure *gradient* with respect to

z is always zero, which guarantees that particle acceleration—and thus, in a scenario with no net flow, the velocity—normal to the surface is zero.

In closing, we note a few consequences of rigid and pressure release surfaces for the transmission of sound in the sea. The linear wave equation predicts the emergence of standing waves in a channel enclosed by two surfaces of either kind described above. If the sea surface and sea bed are modelled as a pressure release surface and a rigid surface, at depths z_s and z_b, respectively, then solutions to the linear wave equation must satisfy:

$$p(t) = -p\left(t + \frac{2k_3(z_s - z_b)}{c}\right). \tag{3.12}$$

Elementary functions that satisfy this equation and the boundary conditions are called *normal modes* of vibration. The first five normal modes are drawn on Figure 3.3C for $k_3 = \pm 1$. The second consequence of surfaces—and perhaps the most relevant to the sonar equations introduced shortly—is that they dictate the spreading geometry. The sound radiating from a spherical source trapped between two parallel, reflective interfaces, e.g., shallow waters or the surface duct, spreads out in cylindrical wavefronts, as Figure 3.3D depicts.

3.1.2 Sources of Sound in the Sea

As we have seen, the signal recorded at a hydrophone is a mixture of pressure waves originating from number of sources over a wide geographical area. This section briefly catalogues some of these sources and their characteristics, working on the principle that a marine vessel is the target for detection and classification, and all other sound is noise.

A large quantity of sound energy is radiated in the course of a ship or submarine being propelled through the water. A modern marine vessel incorporates a wide range of sonorous machinery of specific design and state of repair; these factors determine the sound generated by the vessel at a particular speed and depth. The *acoustic signature* of a target refers to its acoustic spectrum measured at a one metre reference distance (Burdic, 1984) and is highly individual. A passive sonar is therefore able to classify a received target signal by matching it against a list of acoustic signatures stored in a database (although, in practice, the classifier is often a human sonar operator). Our immediate focus concerns the spectral features that constitute the acoustic signature and their physical origin; the techniques for analysing the acoustic signature for the purposes of detection and classification are dealt with in Sections 3.2 and 3.3.

Burdic (1984) identifies four principal sources of sound within the acoustic signature: the propulsion system, auxillary machinery, the propeller and hydrodynamic noise. The propulsion system consists of the engine components responsible for driving the propeller, including shafts, gears, bearings, cylinders, turbines and motors. Each part emits a harmonic spectrum of *machinery lines* with a fundamental corresponding to the rotational frequency. As these components drive the propeller, their frequencies and amplitudes are typically related to the engine speed. Rotatory machines that function independently of the engine, such as on-board electrical generators and air conditioning, generate *auxilliary machine lines*, which are static in frequency and

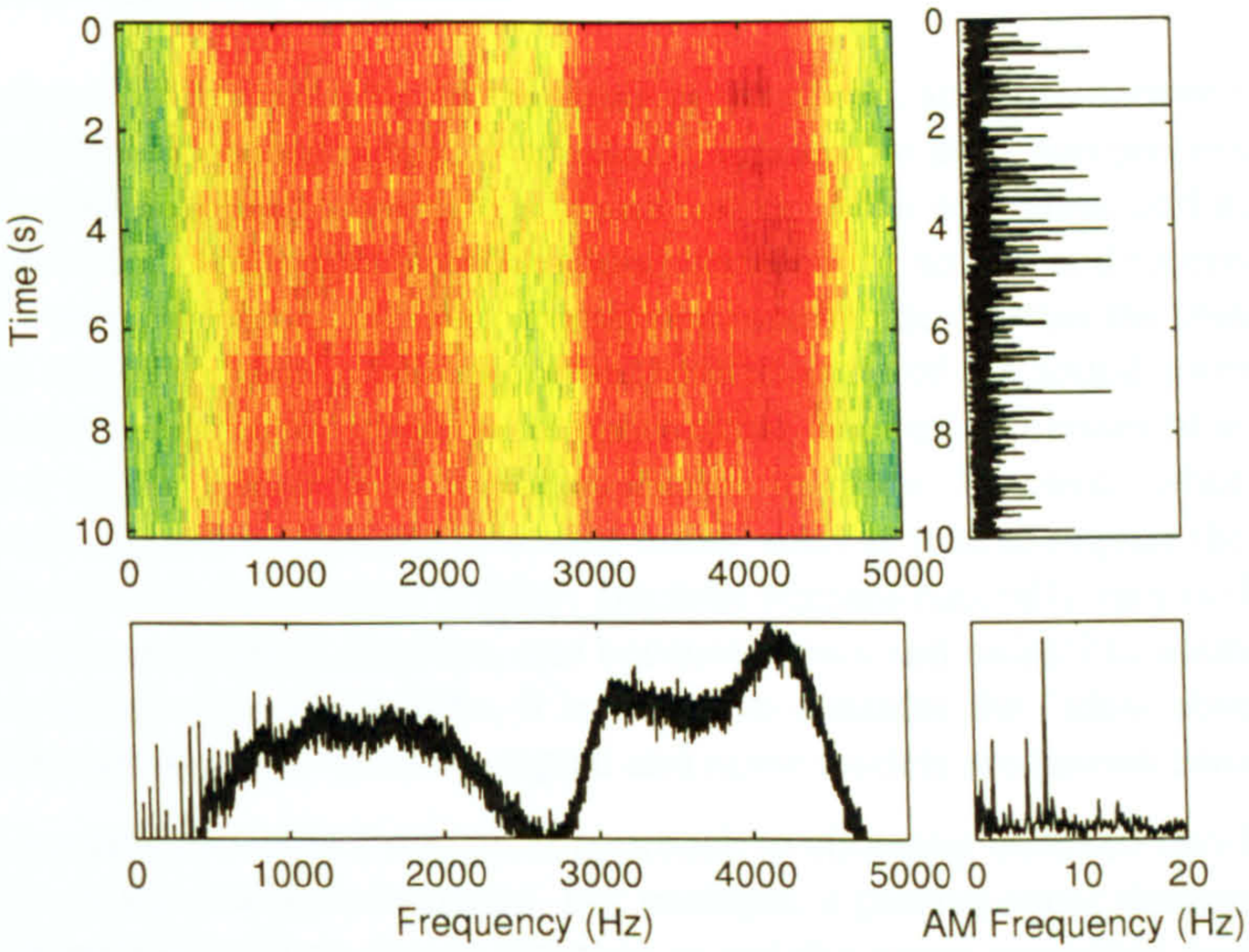


Figure 3.4: Four representations of a merchant vessel signal: log-power waterfall spectrogram (top-left); linear envelope in the time domain (top-right); average power spectrum on a log scale (bottom left); spectral content of envelope (bottom-right).

amplitude. The second component of the acoustic signature, besides the narrowband spectrum, is a continuous, broadband noise spectrum associated with *hydrodynamic* sound, that is, sound generated by the vessel interacting with the surrounding sea water. A major source of hydrodynamic noise is *cavitation*: the formation and collapse of air bubbles at the propeller and on the hull, particularly near the sea surface. Propeller noise is amplitude-modulated at the blade rate.

In both passive and active sonar, the acoustic signatures of nearby vessels or the returning echoes of active pings must be received against a background of ambient noise contributed by the ocean environment as a whole. Ambient noise may be broadly assigned to the following categories: hydrodynamic noise, manmade noise and biological noise (Wenz, 1962). Hydrodynamic noise results from the disturbance of sea water, such as cavitation (bubbles), water droplets, surface waves and turbulence (Wenz, 1962). Hydrodynamic noise in the 100 Hz–10 kHz (mid-frequency) band is correlated with wind speed (Knudsen et al., 1948) and rainfall (Bom, 1969; Scrimger et al., 1987). The principal source of manmade ambient noise is the cumulative noise spectrum arising from distant shipping. In remote ocean regions, shipping noise is confined to frequencies below 100 Hz due to the absorption of high-frequency energy over long distances (Burdic, 1984); in the vicinity of harbours and shipping lanes, the contribution of ocean traffic increases considerably in both bandwidth and intensity (Waite, 1998). Biological noise sources with broadband spectra include snapping shrimp and croakers (Wenz, 1962).

3.1.3 Principles of Optimal Detection

Having reviewed how sound waves originate in the ocean, and summarised the factors that influence how they propagate, it remains to discuss the detection process. The first step of this problem is to nominate a sound source, such as a ship, and account for any channel effects that might alter the sound between the source and receiver, such as those reviewed in Section 3.1.1. The second step is to decide whether the pressure wave recorded at a hydrophone is best explained by a mixture of the sound source with an ambient ocean background, or by ocean noise alone. For certain classes of source, e.g., transients, it is possible to allow a human listener to judge. However, other classes of source, imperceptible to humans in severe noise, such as tonals, require the incoming sound to be converted to a visual display. Because humans naturally vary in their ability to hear sounds and inspect displays, and because tonals and noise-like sounds possess a relatively simple analytical form, it is useful to consider the "ideal observer": the theoretically optimal test, when the signal and noise models are known completely.

A *binary hypothesis test* is a statistical approach to choosing between two hypotheses on the basis of the available evidence. For example, a passive sonar detector performs a binary hypothesis test to decide whether or not the noise-corrupted measurements taken from a receiver array indicate the presence of a target signal. A detection scenario offers two hypotheses: the *null hypothesis* states that only noise has been received; the *alternate hypothesis* states that a mixture of target signal and noise has been received. These hypotheses are random events, which by convention are respectively labelled H_0 and H_1. We can therefore assign a probability to each event, $P(H_0)$ and $P(H_1)$. As one hypothesis is always true to the exclusion of the other, we add the constraint $P(H_0) + P(H_1) = 1$.

If the detector is furnished with no information besides the prior probabilities, the optimum decision rule simply chooses the most probable hypothesis:

$$\text{choose } H_1 \text{ iff } P(H_1) > P(H_0), \text{ otherwise choose } H_0. \tag{3.13}$$

As the detector's decision is uncertain, it is appropriate to characterise the selection of H_0 or H_1 as random events labelled D_0 and D_1. Of course, the ground truth hypothesis, H, and the decision, D, may differ. There are two possible ground truths and two possible decisions, giving a total of four outcomes, which we tabulate below.

	Ground Truth	
Decision	Noise Only (H_0)	Signal and Noise (H_1)
Noise Only (D_0)	True Negative	False Negative
Signal and Noise (D_1)	False Positive	True Positive

Here, *negative* and *positive* refer to the detector's decision, i.e., that signal is absent or present, respectively; *true* and *false* indicate whether the decision is *correct* (not the ground truth). The expressions *detection* and *false alarm* are synonymous with true positive and false positive, respectively. These four random events are associated with the conditional probability $P(D_j \mid H_k)$, i.e., the probability that the detector chooses D_j, given that H_k is true. The decision is correct if and only if $j = k$.

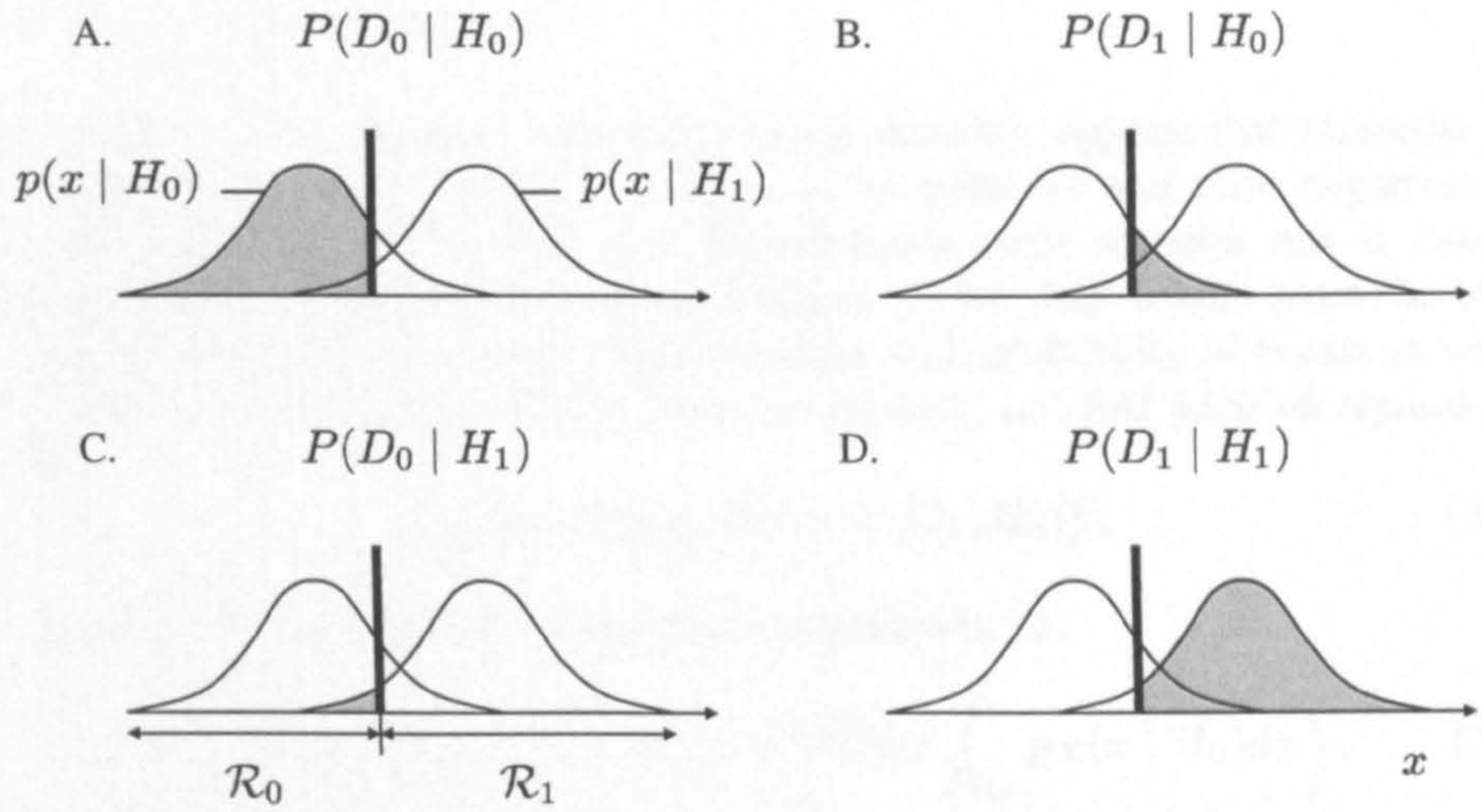


Figure 3.5: A threshold (thick black line) on x divides the x-axis into two decision regions $\mathcal{R}_0$ and $\mathcal{R}_1$, indicated in plot (C). The shaded gray areas represent the following conditional probabilities: A) correct dismissal; B) false alarm; C) false dismissal; D) correct detection.

Naturally, for a detector to be of practical use, it must incorporate observations into the decision process. The input supplied to the detector is typically some sort of summary value distilled from the raw data called a *test statistic*. For instance, in passive sonar, "raw data" might refer to the time-varying voltage output of a receiver and the test statistic its root-mean-square. Section 3.2 describes the use of test statistics based on power, and Section 3.3 explores the possibility of using a test statistic based on timing. Here, we designate a general test statistic x and assume it is governed by a random variable X. Using these definitions, any binary detector may be considered a deterministic function that maps every element in the domain of the test statistic (e.g., the real numbers) to a decision.

The domain is split into two *decision regions*, $\mathcal{R}_0$ and $\mathcal{R}_1$, which respectively contain the elements that map to D_0 and D_1. Figure 3.5 shows the two distributions X assumes, given H_0 or H_1, and two decision regions formed by partitioning the number line with a single cut. The detection outcomes listed in the four table cells above, i.e., true/false positive/negative, can be found by integrating the conditional p.d.f. of X in the appropriate decision region:

$$P(D_j \mid H_k) \equiv \int_{\mathcal{R}_j} p_X(x \mid H_k) dx. \tag{3.14}$$

The final step, having defined the relevant symbols and quantities, is the adjustment of $\mathcal{R}_0$ and $\mathcal{R}_1$ to maximise the performance of the detector according to some criterion.

Minimum Error Decision Criterion

The *minimum error decision rule* employs the decision regions that minimise the probability of any kind of error occurring; false positives and false negatives are presumed to carry equal weight. As the minimum error decision rule is easy to obtain and used extensively throughout Chapter 4, we shall briefly trace the steps in its derivation here. We wish to minimise the total probability of events in which the detector decision and the ground truth are opposite, i.e., find decision regions that satisfy

$$\arg\min_{\mathcal{R}_1} \{P(D_0, H_1) + P(D_1, H_0)\}. \tag{3.15}$$

Using (3.14), we can write the probabilities as integrals, i.e.,

$$\arg\min_{\mathcal{R}_1} \left\{ P(H_1) \int_{\mathcal{R}_0} p_X(x \mid H_1) dx + P(H_0) \int_{\mathcal{R}_1} p_X(x \mid H_0) dx \right\}. \tag{3.16}$$

Evidently, any x will contribute to *exactly* one of the two additive terms in (3.16), depending on whether it is a member of $\mathcal{R}_0$ or $\mathcal{R}_1$. As we are free to choose these regions, we can minimise the braced expression by choosing $\mathcal{R}_1$ (and implicitly, $\mathcal{R}_0$) so that x always contributes to the lesser of the two terms, i.e.,

$$\mathcal{R}_1 = \{x : P(H_1) p_X(x \mid H_1) > P(H_0) p_X(x \mid H_0)\}. \tag{3.17}$$

From (3.17), we arrive at the minimum error decision rule:

$$\text{choose } H_1 \text{ iff } \frac{p_X(x \mid H_1)}{p_X(x \mid H_0)} > \frac{P(H_0)}{1 - P(H_0)}, \text{ otherwise choose } H_0. \tag{3.18}$$

Assuming that the prior probabilities and p.d.f.s accurately describe the distribution of the data, the decision rule (3.18) is optimal in the sense that no other detector exists that commits fewer errors. Note that in the absence of any test statistic, the minimum error decision rule reduces to (3.13).

Neyman-Pearson Decision Criterion

Often, rather than minimising the overall probability of error, it is preferable to specify a fixed false positive probability, whilst simultaneously minimising the probability of false negatives. This is particularly useful in that it allows one to partially quantify the performance of an envelope or amplitude detector, even when the signal-and-noise (H_1) distribution is unknown (Burdic, 1984; Whalen, 1971). As before, we choose the decision region $\mathcal{R}_1$, and its complement $\mathcal{R}_0$, to satisfy

$$\arg\min_{\mathcal{R}_1} \{P(D_0, H_1) + P(D_1, H_0)\}, \tag{3.19}$$

only this time include the constraint

$$P(D_1 \mid H_0) = p_{fa}, \tag{3.20}$$

where p_{fa} is a constant. Simultaneously solving (3.19) and (3.20) using the method of Lagrange multipliers (Whalen, 1971, page 133) obtains decision regions that describe

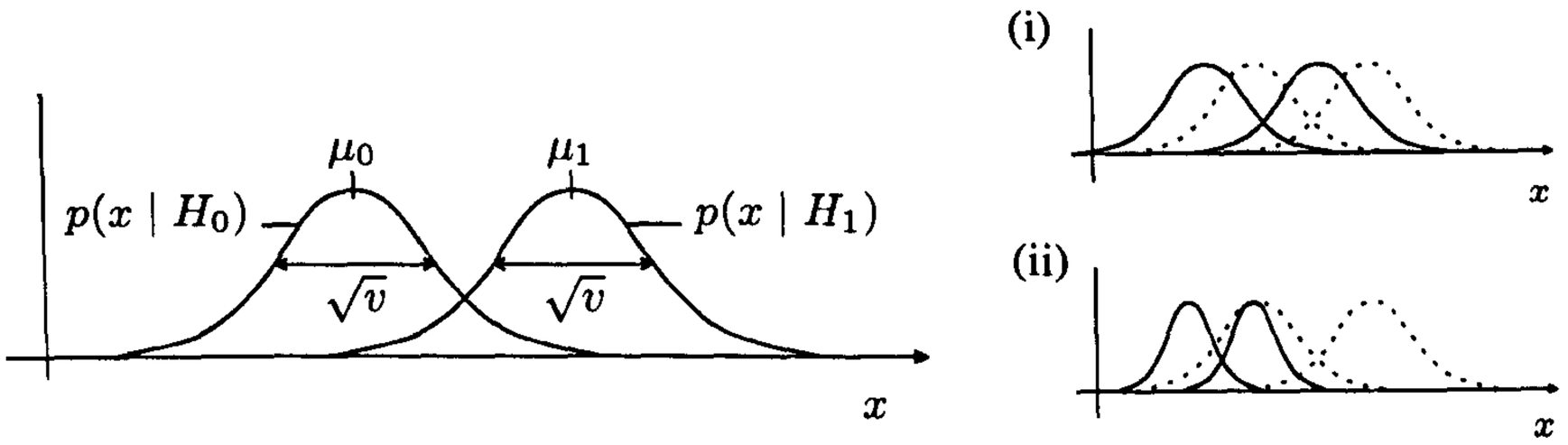


Figure 3.6: A sketch of two overlapping normal probability density functions with means μ_0 and μ_1, and identical variances, $v_0 \equiv v_1 \equiv v$. Note that neither (i) shifting nor (ii) dilating the distributions on the x-axis affects the separability of the classes.

a generalised version of the minimum error rule, called the *Neyman-Pearson decision rule*:

$$\text{choose } H_1 \text{ iff } \frac{p_X(x \mid H_1)}{p_X(x \mid H_0)} > \eta, \text{ otherwise choose } H_0. \tag{3.21}$$

The parameter η on the right-hand side of the inequality in (3.21) can be thought of as a "sliding threshold", which adjusts the sensitivity of the detector. At this juncture, it is appropriate to introduce ROC curves.

ROC Curves and the Detection Index

A *receiver operating characteristic* (ROC) curve provides a means of visualising the trade-off between detection and false alarm probability as the threshold η is varied. Specifically, a ROC curve is defined by the locus

$$ROC = \{(p_{fa}, p_d) : p_{fa} = P(D_1 \mid H_0; \eta), p_d = P(D_1 \mid H_1; \eta), \forall \eta > 0\} \tag{3.22}$$

and is plotted on a pair axes, each axis ranging from 0 to 1. Stated another way, a ROC curve is an explicit function mapping the probability of false alarm to the probability of detection (Burdic, 1984). We shall restrict the focus of this section to the ROC curves for Gaussian conditional p.d.f.s.

The need to decide from which of two Gaussian distributions a sample has been drawn arises in many detection and classification applications, including passive sonar. If the variances of the two distributions are identical (or near-identical), then the inherent difficulty of the problem may be characterised by three parameters: the two means, μ_0 and μ_1, and the variance, v, as shown in Figure 3.6. For instance, one might express the probability of false alarm as a function, $p_{fa}(\mu_0, \mu_1, v)$. Translating the conditional p.d.f.s along the x-axis by the same amount (i.e., shifting both means) does not affect the separability; hence we can equivalently write $p_{fa}(0, \mu_1 - \mu_0, v)$. Similarly, dividing the test statistic by a constant leaves the separability unaffected, and so we can obtain an equivalent parameterisation, $p_{fa}(0, 1, v/(\mu_1 - \mu_0)^2)$, or simply, $p_{fa}(d^{-1})$, where

$$d = \frac{(\mu_1 - \mu_0)^2}{v}. \tag{3.23}$$

The normalised parameter d is referred to as a the *detection index* (Waite, 1998; Dawe, 1997). Two detection tasks with identical detection indices are of the same intrinsic difficulty; for example, halving the separation between the distribution means creates a detection problem equivalent to that obtained by quadrupling the distribution variances. ROC curves for various detection indices are plotted in Figure 3.7.

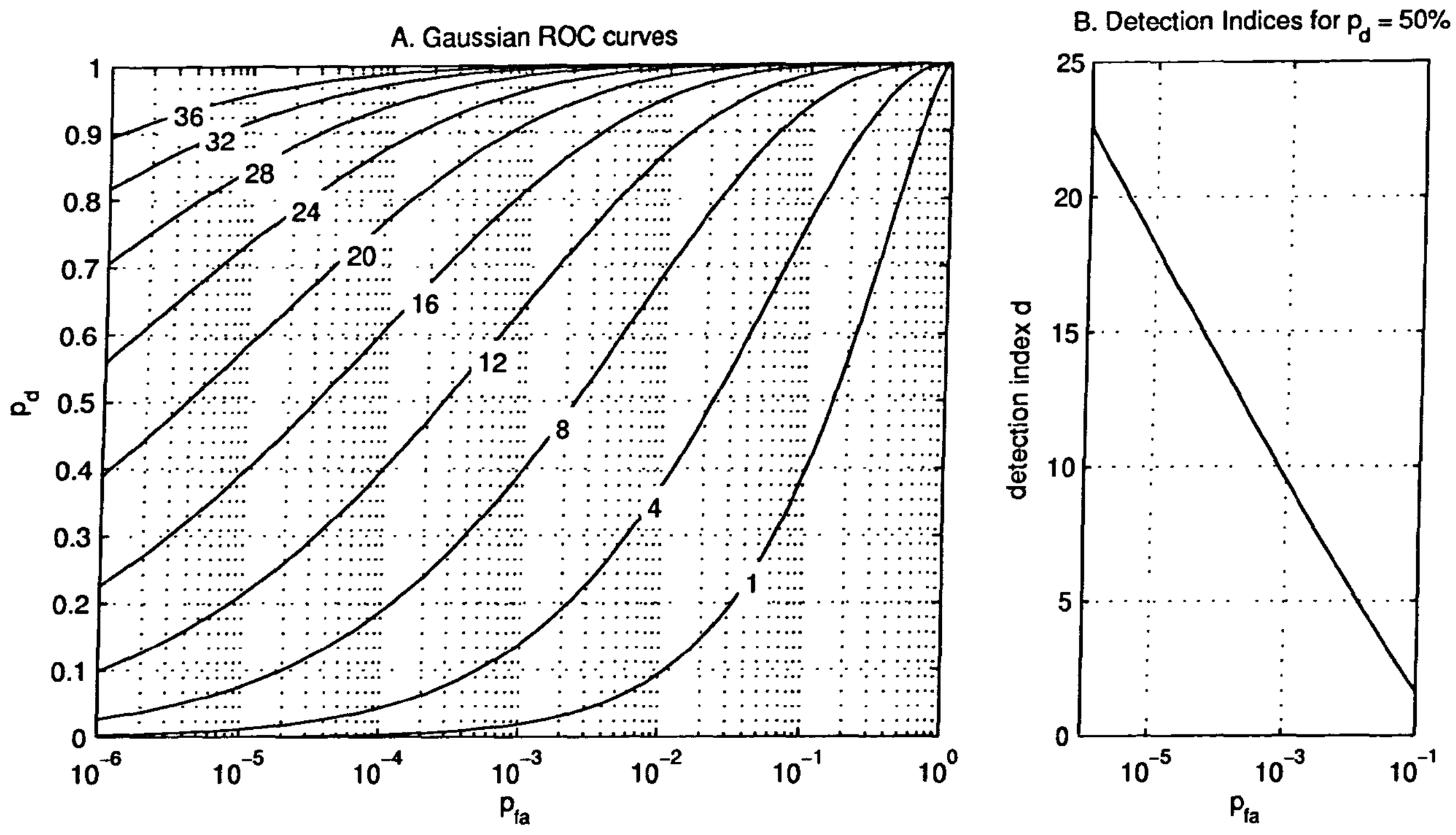


Figure 3.7: ROC curves for Gaussian statistics produced using MATLAB. A) a family of ROC curves plotting the probability of detection as a function of probability of false alarm. The detection index is marked on each curve. B) the detection index required to satisfy a probability of false alarm (on the abscissa) and a fixed detection probability of 50%. For an extensive collection of ROC curves, see Whalen (1971).

3.2 Power-based Detection

3.2.1 Passive Broadband Detector

A conventional passive broadband detector compares the mean-square of the signal envelope to a threshold in order to determine whether a signal is present. The receiver comprises four stages, shown as a block diagram in Figure 3.8. The first three stages take the form of an envelope detector sandwiched between two linear filters. The filter preceding the envelope detector is called *predetection filter* and has impulse response $h_a(t)$; the filter following the envelope detector is called the *postdetection filter* and has impulse response $h_b(t)$. The fourth stage is a threshold chosen according to suit a detection criterion, e.g., minimum error or Neyman-Pearson.

Depending whether the input signal $g(t)$ is a noise-only signal (H_0) or a mixture of signal and noise (H_1), the power spectral density at the the output of the predetection filter is one of the following:

$$\mathscr{S}_0(\omega) = |\mathscr{H}_a(\omega)|^2 \mathscr{S}_n(\omega) \tag{3.24}$$

$$\mathscr{S}_1(\omega) = |\mathscr{H}_a(\omega)|^2 \left[\mathscr{S}_s(\omega) + \mathscr{S}_n(\omega)\right], \tag{3.25}$$

where $\mathscr{S}_n$ and $\mathscr{S}_s$ respectively denote the p.s.d.s of the signal and noise processes at the receiver input. The output of the predetection filter for hypothesis H_j is a Gaussian process with variance (i.e., power),

$$\sigma_j^2(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathscr{S}_j(\omega) d\omega. \tag{3.26}$$

The instantaneous output of the squared-envelope block is governed by an exponential distribution with mean $\mu = 2\sigma^2$ and variance $v = 4\sigma^4$, if the input is a zero-mean stationary Gaussian process with variance σ^2. According to the central limit theorem (Peebles, 1993), the average of n independent samples of the envelope follows an

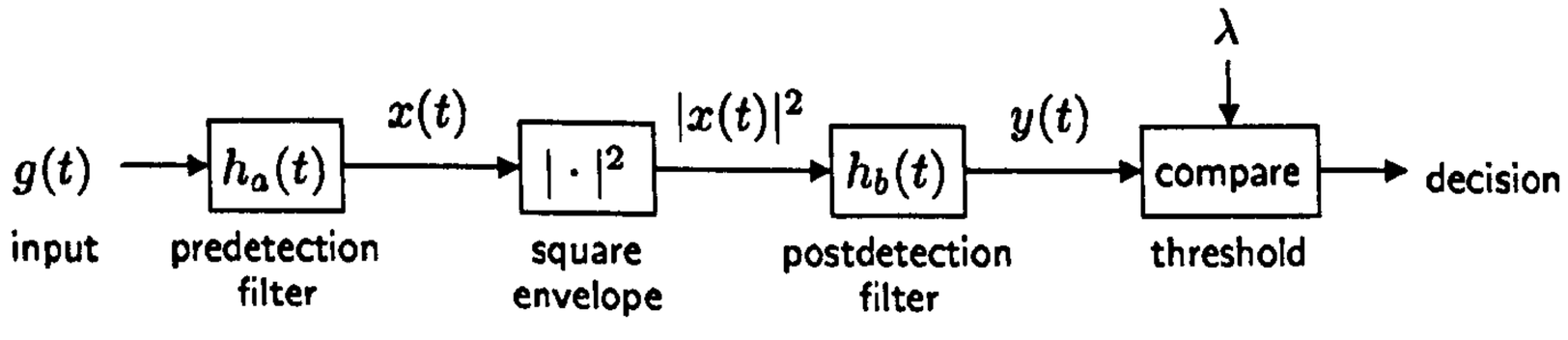


Figure 3.8: Passive broadband detector block diagram. The input signal $g(t)$ is filtered and the squared-envelope is measured at the filter output; a second linear filter then averages the output of the envelope detector over a sufficiently long time period, so that the signal $y(t)$ conforms to a Gaussian process. The final block is a decision rule based on Gaussian probability density functions, which compares a sample of y to the threshold λ.

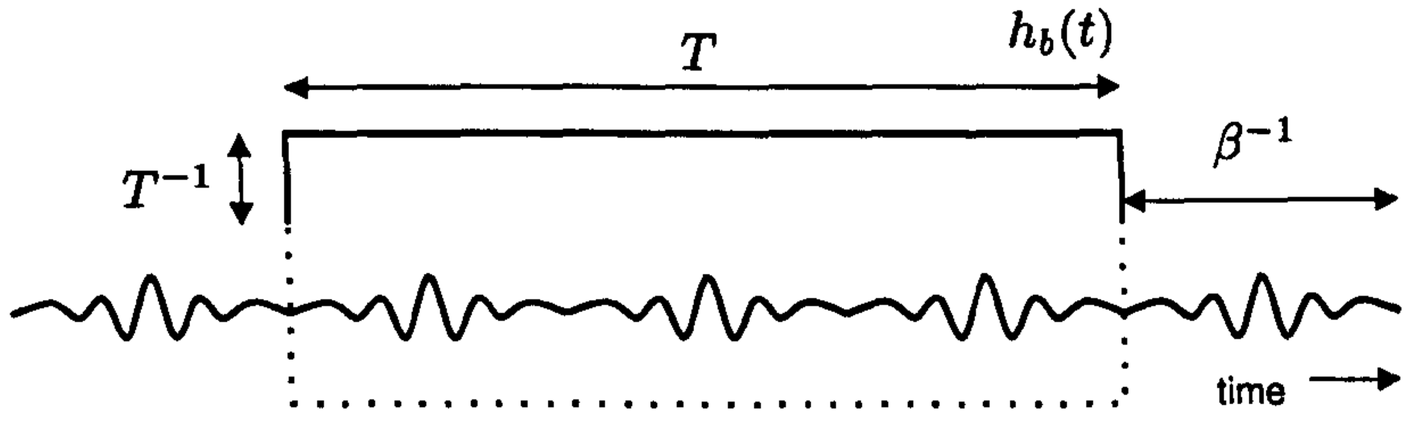


Figure 3.9: Time-bandwidth product. If a noise process is conceived as arising from white noise samples convolved with an impulse response of duration β^{-1}, then a time window T seconds long provides at least βT independent Gaussian samples.

approximately Gaussian distribution with

$$\text{mean} = \mu = 2\sigma^2, \text{ variance} = v = \frac{4\sigma^4}{n},$$

provided that n is large. It is the role of the postdetection filter to average independent samples of the squared-envelope in order to obtain Gaussian statistics. There are various possible choices of impulse response $h_b(t)$ that may achieve this; perhaps the most obvious is a rectangular pulse of unit area, i.e., T seconds in duration and $1/T$ in amplitude. We can gain an idea of how many independent samples are combined in the decision rule by considering the effective bandwidth of the noise, β Hz. Using the rule-of-thumb that the impulse response duration is reciprocally related to the bandwidth, the number of independent samples produced in time T is βT. The postdetection window shown in Figure 3.9, for instance, is long enough to contain three non-overlapping copies of the impulse response, so the time-bandwidth product is $\beta T = 3$.

If $\sigma_1^2 \approx \sigma_0^2$, i.e., the signal-to-noise ratio after predetection filtering is low, then a good approximation to the detection index is found by placing these parameters into (3.23), yielding

$$d \approx \frac{\beta T(2\sigma_1^2 - 2\sigma_0^2)^2}{4\sigma_0^4} = \beta T \left(\frac{\sigma_s^2}{\sigma_n^2}\right)^2. \tag{3.27}$$

The detection index can be used in connection with the ROC curves in the preceding section to obtain suitable trade-off between false alarm rate and detection probability. It is also instructive to note that the detection index is a product of two meaningful factors: i) the signal-to noise ratio at the output of the predetection filter and ii) the number of independent samples that make up the test statistic—increasing either (i) or (ii) will improve the performance of the detector.

3.2.2 Passive Broadband Sonar Equations

In this section, we shall briefly demonstrate how the *passive sonar equation* can be used to predict some aspect of a sonar's performance under specific conditions. The sonar equation, in its most basic form, is a sum of decibel quantities:

$$\mathsf{SE} = \mathsf{S} - \mathsf{N} - \mathsf{DT}. \tag{3.28}$$

S and N respectively denote the signal and noise level following beamforming and predetection filtering; consequently, S − N is the signal-to-noise ratio in the receiver. The terms SE and DT correspond to *signal excess* and *detection threshold*: the signal excess relates the extent (in dBs) to which the SNR exceeds that required for some prestated performance, characterised by the detection threshold; when SE = DT there is no signal excess. SE and DT are reciprocal quantities: raising the detection threshold lowers the signal excess and *vice versa*.

The signal level at the receiver, S, may be expanded further into two terms: the source level (SL) and the propagation loss (PL). To illustrate this principle, we can imagine that a target is radiating sound energy with intensity SL dB at one metre distance, and due to boundary effects in shallow water, the sound spreads away from the target in concentric cylinders. Applying the formula for the sound intensity on the surface of a cylinder of height h in (3.9), the source level expressed in decibels (with respect to a reference intensity I_{ref}) is given by

$$\mathsf{SL} = 10 \log_{10} \frac{\Pi_{cyl}/(2\pi h)}{I_{ref}}, \text{ dB}. \tag{3.29}$$

If the signal propagates over a range R metres before reaching the array, then the intensity at the received wavefront is

$$\mathsf{S} = 10 \log_{10} \frac{\Pi_{cyl}/(2\pi \cdot Rh)}{I_{ref}}, \text{ dB}. \tag{3.30}$$

The propagation loss is defined as the ratio of source intensity to received intensity; thus, combining (3.29) and (3.30), we may express this relationship as a linear relationship between levels:

$$\mathsf{PL} = \mathsf{SL} - \mathsf{S} = 10 \log_{10} R, \text{ dB}. \tag{3.31}$$

The detection threshold is defined as the signal-to-noise ratio that results in prespecified detection and false alarm probabilities. If the test statistic is the averaged output of a squared-envelope detector, and the time-bandwidth product $\beta T \gg 1$, then the detection threshold DT shares the following relationship with the detection index d:

$$\mathsf{DT} = 5 \log_{10} d - 5 \log_{10} \beta T, \tag{3.32}$$

found by rearranging (3.27). Replacing (3.31) and (3.32) in (3.28) gives an expanded form of the sonar equation for broadband detection,

$$\mathsf{SE} = \mathsf{SL} - \mathsf{PL} - \mathsf{N} - 5 \log_{10} d + 5 \log_{10} \beta T. \tag{3.33}$$

A brief example will demonstrate the utility of this sonar equation. Suppose the target radiates noise at 140 dB, eight kilometres from the receiver. What ambient noise level

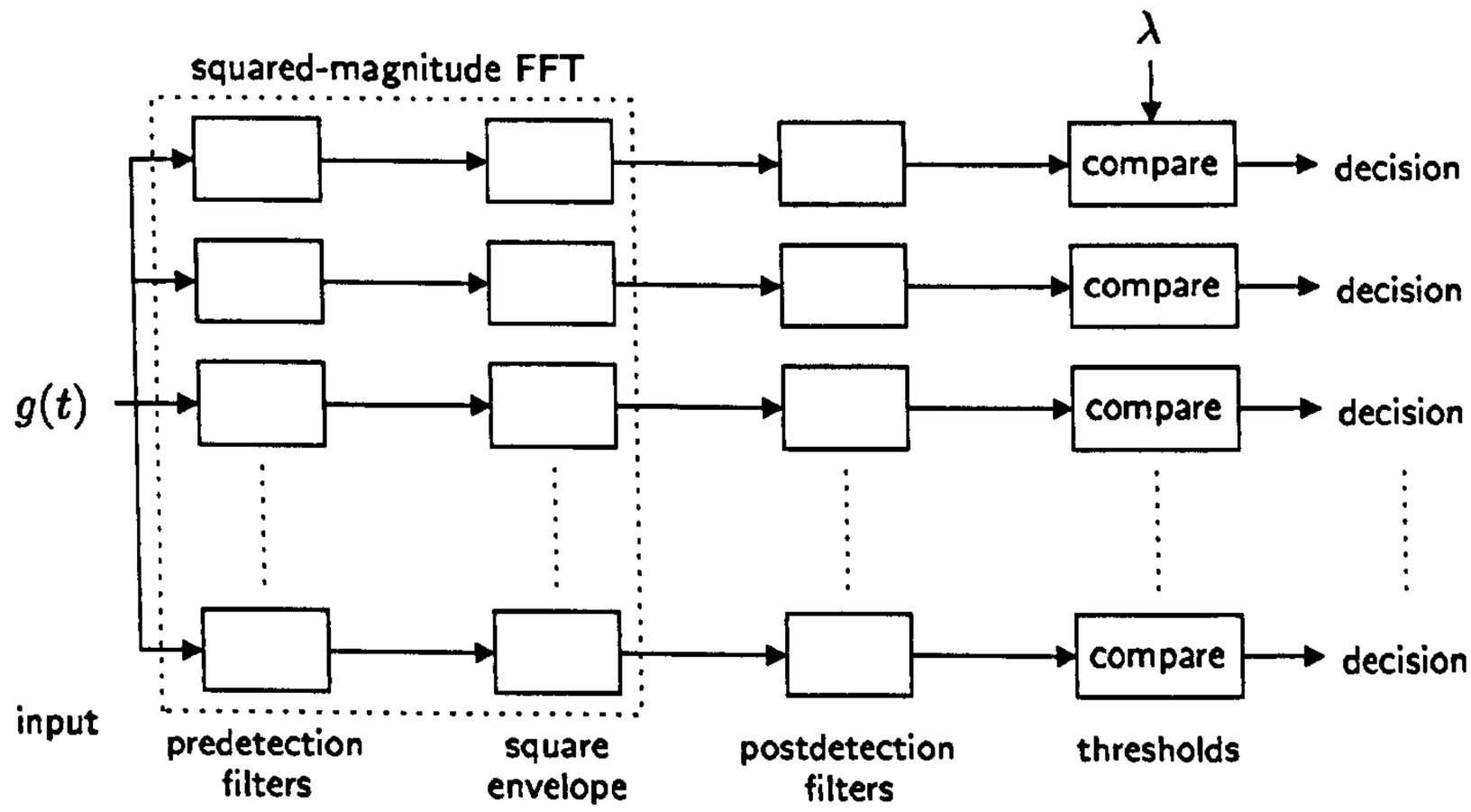


Figure 3.10: Passive narrowband detector block diagram. Each parallel pathway in the detector functions in the same way as a broadband detector (*cf.* Figure 3.8). The first two steps are performed for all channels by a squared magnitude fast Fourier transform.

is required to secure a probability of detection of 50% and a probability of false alarm of 0.1%, if samples are averaged over 10 seconds in a 1 kHz bandwidth? First of all,

$$\mathsf{PL} = 10\log_{10} 8000 = 39 \text{ dB, [from (3.31)]}$$

$$\beta T = 10 \times 1000 = 10,000.$$

Then, reading from the ROC curve in Figure 3.7B, the detection index corresponding to the desired performance is $d = 9$. Placing these values into the sonar equation leaves an expression in N,

$$0 = 140 - 39 - \mathsf{N} - 4.77 + 20;$$

therefore, the ambient noise level must not exceed 116 dB.

3.2.3 Passive Narrowband Sonar Equations

Passive narrowband sonar is used to detect tonals and operates on the same principle as broadband detection—i.e., a hypothesis test based on Gaussian statistics—with the exception that the incoming signal is prefiltered into many narrow bands. In practice, the signal is sampled at a rate f_s Hz, and the narrowband filtering is accomplished in parallel by a fast Fourier transform (FFT) processor. Concerning performance analysis, it is helpful to note that the FFT magnitude samples are almost identical to the envelope that would be sampled across a bank of N filters with finite impulse responses

$$h_s[n] = \begin{cases} w[N-n-1]\cos 2\pi s\frac{n+1}{N} & 0 \le n \le N-1 \\ 0 & \text{otherwise,} \end{cases} \tag{3.34}$$

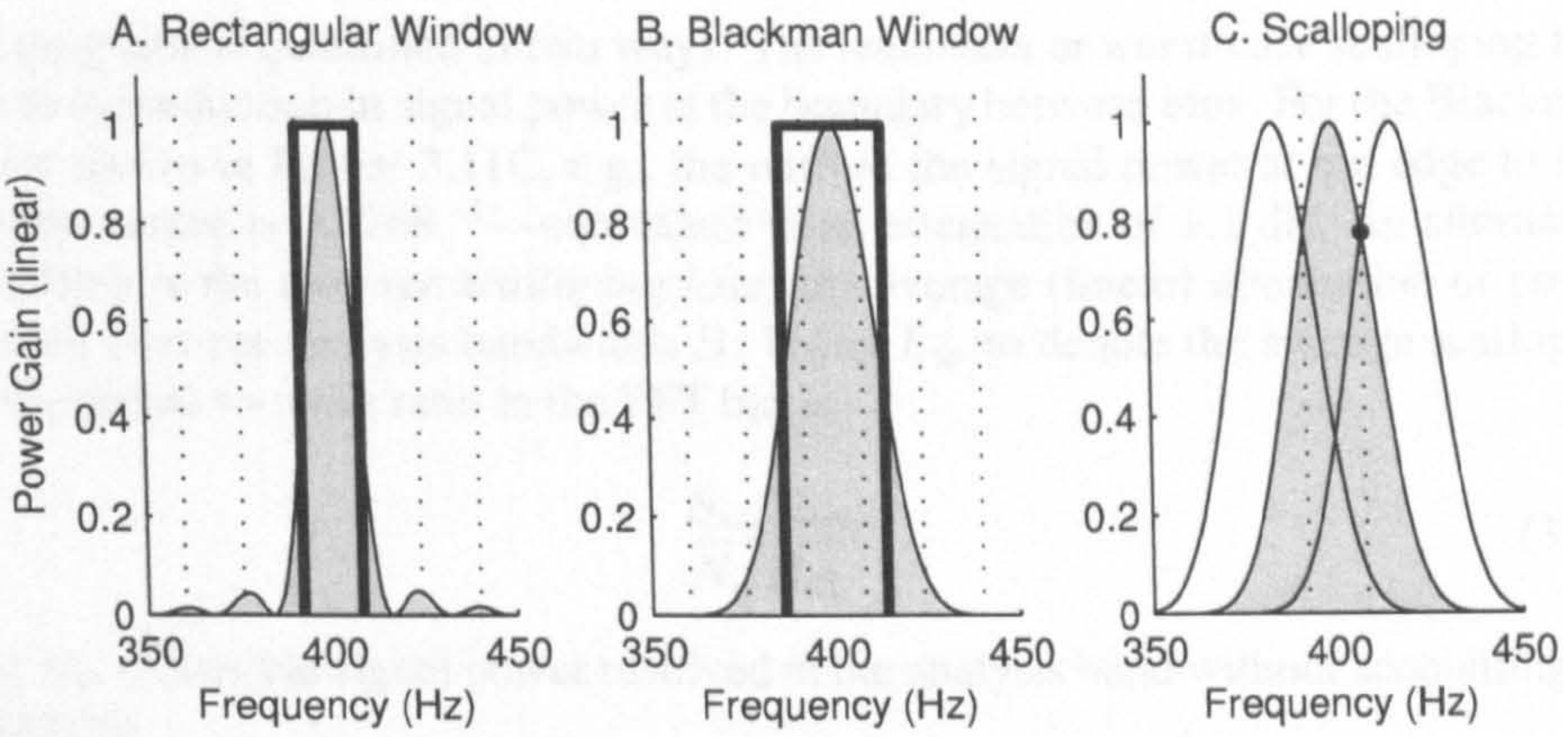


Figure 3.11: The effect of windowing upon the FFT bin squared magnitude response. Analysis bins are demarcated by dotted lines. A) a rectangular window causes spectral leakage in the frequency domain; B) the Blackman window offers good sidelobe attenuation and a broader passband. The effective noise bandwidth for each filters is superimposed in a heavy stroke. C) worst-case scalloping loss occurs halfway between bins, as indicated by the marker, where signal power is scaled by a factor of 0.7768 or -1.10 dB (Harris, 1978).

where s is an integer indexing the bin with centre frequency sf_s/N Hz, and $w[\cdot]$ is the window function. The N-point FFT has an analysis binwidth given by

$$B = f_s/N, \text{ Hz} \tag{3.35}$$

and a baseband bandwidth equal to $\frac{1}{2}f_s$ Hz. In a narrowband FFT, B is by definition small, so it is generally assumed that the noise power spectral density across the filter is sufficiently smooth that it can be modelled as a constant, N_w, referring to the power resolved in a 1 Hz bandwidth. The average noise power contributed to an FFT cell is then given by $N_w B$.

The frequency response of a 1024-point FFT bin is shown in Figure 3.11A. The sample rate is 16384 Hz, the bin centre frequency is 400 Hz, the binwidth is $B = 16$ Hz, and $w[\cdot]$ is a rectangular window. The sharp edges of the rectangular window result in *spectral leakage*, manifested as sidelobe responses at ± 23 Hz, ± 39 Hz, etc. The response of the FFT bin can be smoothed out by selecting a tapered window for $w[\cdot]$, prior to performing the FFT, at the expense of broadening the mainlobe. Harris (1978) has documented the key properties of a variety window functions, two of which are relevant here. The first is *effective noise bandwidth*: "the width of a rectangular filter with the same peak power gain that would accumulate the same noise power" (Harris, 1978), which we designate B_{eff}. For example, the Blackman window, whose frequency domain squared magnitude response is plotted in Figure 3.11B, has a bandwidth 1.73 times wider than the analysis bandwidth, i.e., approximately 27.68 Hz. Naturally, the adjustment for effective noise bandwidth implies a decrease in SNR.

The second property of the window function is *scalloping loss*, which is defined as the loss in signal power (or SNR) due to a sinusoid being resolved between FFT bins.

Scalloping loss is quantified in two ways. The *maximum* or *worst-case* scalloping loss refers to the reduction in signal power at the boundary between bins. For the Blackman window shown in Figure 3.11C, e.g., the ratio of the signal power at bin edge to that at the bin centre is 0.7768^{-1}—equivalent to an attenuation of 1.1 dB. An alternative formulation is the *average scalloping loss*: the average (linear) attenuation of power measured over the analysis bandwidth B. Using L_{as} to denote the average scalloping loss, the signal-to-noise ratio in the FFT bin is

$$\frac{S_w/L_{\text{as}}}{N_w B_{\text{eff}}}, \tag{3.36}$$

where S_w relates the signal power resolved in the analysis band without accounting for windowing.

Statistical detection using the FFT follows the same principle as a broadband scheme described above: many consecutive, independent FFT bins are averaged in order to obtain Gaussian statistics. In a narrowband context, the number of FFT blocks that are averaged incoherently is referred to as the *integration factor* (IF) and is a quantity analogous to the time-bandwidth product in broadband sonar. If only Gaussian noise is present, then the mean of IF $\gg 1$ squared magnitude samples of an FFT bin follows a Gaussian distribution with mean and variance respectively given by

$$\mu_0 = B_{\text{eff}} N_w \tag{3.37}$$

$$v = (B_{\text{eff}} N_w)^2/\text{IF}. \tag{3.38}$$

$$\mu_1 = S_w/L_{\text{as}} + B_{\text{eff}} N_w. \tag{3.39}$$

Consequently, under the assumption of a low signal-to-noise ratio, the detection index is found from (3.23) to be

$$d = \frac{\text{IF}}{L_{\text{as}}^2 B_{\text{eff}}^2} \left(\frac{S_w}{N_w}\right)^2, \tag{3.40}$$

from which the expression for the detection threshold follows,

$$\mathsf{DT} = 5\log_{10} d - 5\log_{10} \text{IF} + 10\log_{10} B_{\text{eff}} + 10\log_{10} L_{\text{as}}. \tag{3.41}$$

The passive narrowband sonar equation is obtained by replacing DT in (3.28).

$$\mathsf{SE} = \mathsf{SL}_w - \mathsf{PL} - \mathsf{N}_w - 5\log_{10} d + 5\log_{10} \text{IF} - 10\log_{10} B_{\text{eff}} - 10\log_{10} L_{\text{as}}. \tag{3.42}$$

For a more thorough-going discussion of the quantities that appear in the narrowband sonar equation, see Waite (1998), Dawe (1997) or Burdic (1984, Chapters 13 and 14).

3.3 Timing-based Detection

Chapters 1 and 2 introduced auditory-motivated and temporal signal representations from a physiological and signal processing perspective. Several examples of temporal representations were also reviewed: the zero crossing with peak amplitudes (ZCPA), the ensemble interval histogram (EIH), the generalised synchrony detector (GSD), synchrony strands and autocorrelation. Of these candidates, only the ZCPA shall be considered for adaptation to sonar applications. The rationale for this decision is twofold.

First, of all the models, the ZCPA bears the closest resemblance to more contemporary time-frequency representations, such as the sparse time-frequency representation (Gardner and Magnasco, 2006) and the reassigned spectrogram (Fulop and Fitz, 2006; Kodera et al., 1976, 1978). Note that both of these representations make use of the instantaneous frequency, of which a zero crossing interval is simply a measurement.

Second, the ZCPA incorporates many features of the other models. Kim et al. (1999) explains how the ZCPA was developed from, and is in many regards superior to, the EIH. (See also Section 2.2.5.) A synchrony strand (Cooke, 1991/1993) is formed when a block of auditory filters phase-locks to a single instantaneous frequency. This idea is analogous to peak tracking in the ZCPA, as a dominant component causes identical zero crossing intervals to be received across groups of channels, which then contribute to a single peak in the ZCPA histogram.

3.3.1 ZCPA with Auditory-like Parameters

The general purpose of this section is to address whether the zero crossings with peak amplitudes algorithm can be adapted to suit narrowband sonar applications and, if so, to identify the kinds of adjustment that are required. As a starting point, we will compute the ZCPA for the first four seconds of the sonar signal shown in Figure 3.4, using the same set of parameters that were used to produce the speech ZCPA in Figure 2.5. These parameters are tabulated in Table 3.1 and are fairly typical of those used in auditory models. The resulting ZCPA time-frequency display is shown in Figure 3.12 in a waterfall format[1].

The ZCPA provides a broadband analysis of the sonar signal but is unable to resolve narrowband features such as tonals. (That the signal contains low-frequency tonals can be verified from the summary spectrogram in Figure 3.4, and later studies in this section confirm this.) Although intermittent spectral lines are discernible at low frequencies, suggesting the presence of some tonal structure, the ZCPA retains too little detail to conclude anything further, and the time-averaged summary ZCPA does not exhibit any well-defined peaks.

The inability of the ZCPA to resolve narrowband features can be traced to a number of factors, which are summarised in Table 3.2 and described now in detail. The first

[1] that is, frequency along the abscissa and time running down the ordinate.

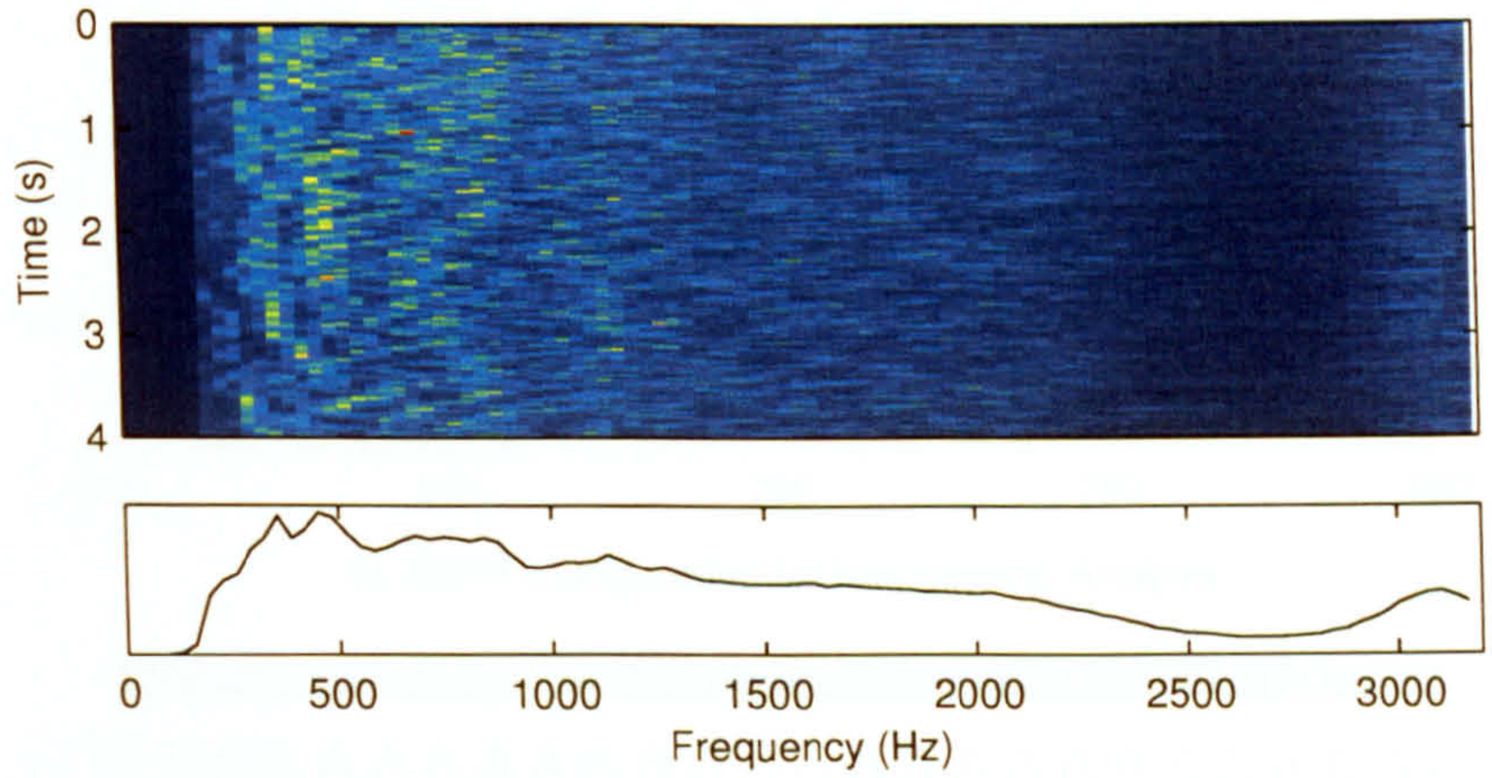


Figure 3.12: A waterfall ZCPA for four seconds of a recorded sonar signal. The lower plot shows the summary pseudo-spectrum computed by averaging the ZCPA over time.

and most obvious factor is the relatively wide bandwidths of the gammatone filters. The 3 dB bandwidths range from approximately 41 Hz to 328 Hz, and, in this regard, the frequency-resolving power of the ZCPA faces the same limitations as a wideband Fourier analysis, specifically, a low post-analysis signal-to-noise ratio and the inability to separate closely-spaced components.

Unlike an FFT, however, the frequency resolution of the ZCPA is enhanced by the fine structure analysis carried out by the zero crossing detection and histogram compilation. Following the analysis filterbank, the first trade-off is the number of intervals that are extracted from each channel in order to form the histogram. A longer window adds more intervals to the ZCPA histogram, resulting in a smoother profile; but a shorter window measures frequency on a shorter time scale and is therefore better-suited to the analysis of non-stationary signals.

AUDITORY-LIKE PARAMETERS

Parameter	**Value**
channel range	200 Hz – 3200 Hz
channel resolution	85 gammatones, ERB-spaced
bandwidths	ERB
peak compression	$\log(x+1)$
interval/peak window	20 most recent intervals
histogram range	0 Hz – 3200 Hz
histogram resolution	100 bins, linearly-spaced
record ZCPA at intervals	5 ms

Table 3.1: Auditory-like ZCPA parameter set, chosen to reflect values typically used in auditory modelling studies (Kim et al., 1999; Ghitza, 1988).

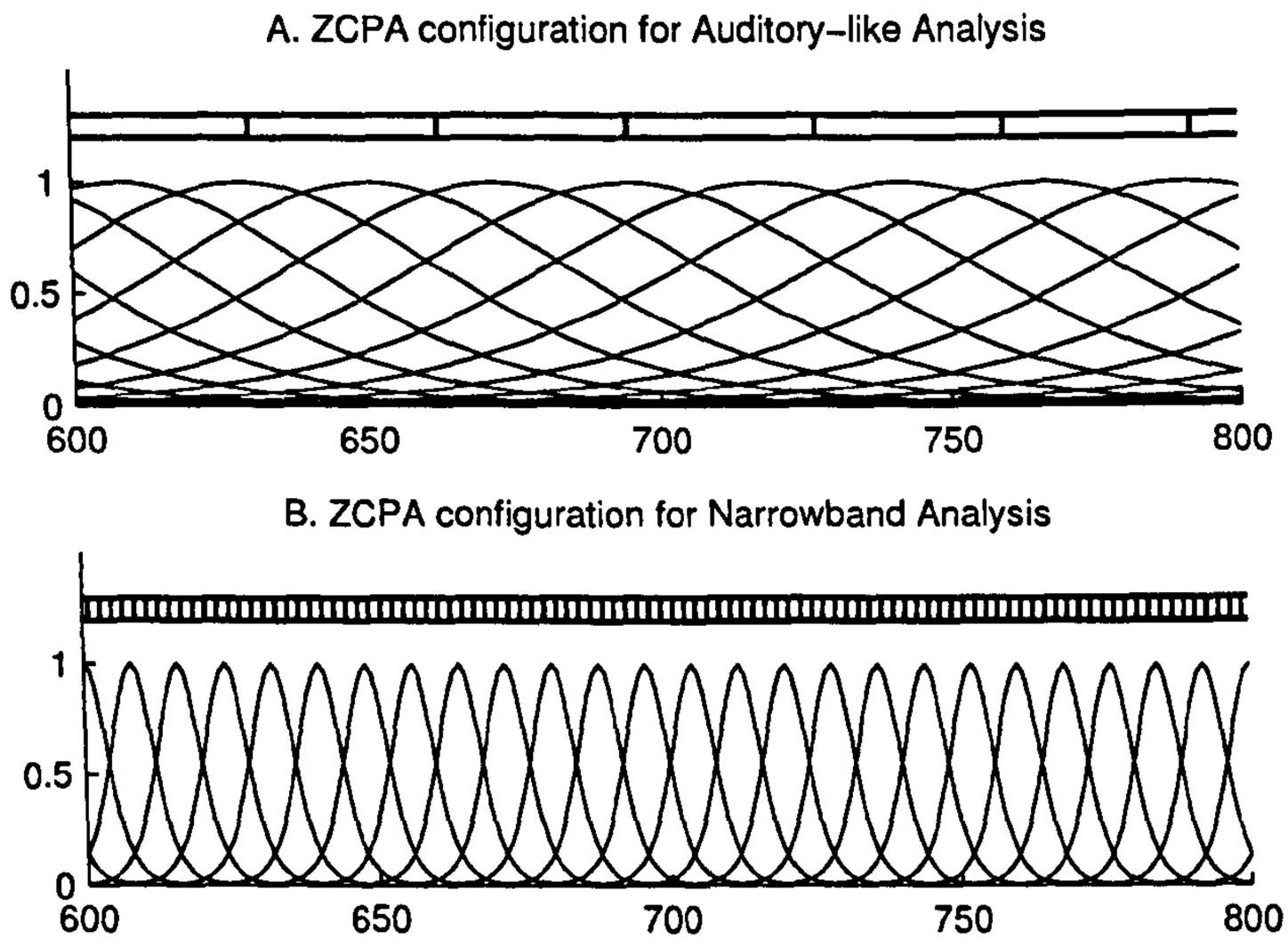


Figure 3.13: ZCPA analysis resolution. A) the squared magnitude response of the filterbank that results from using the parameter set in Table 3.1. The histogram bins are drawn above this. Note that, due to ERB spacing, the filters are somewhat narrower and more tightly-packed at lower frequencies, and broader and sparsely-distributed at higher frequencies. B) the squared magnitude response of the filterbank and histogram configuration that result from using the parameter set in Table 3.3.

At each time-step in the ZCPA, once the intervals have been extracted, there remains the question of how to assign them to the histogram. If the histogram bin widths are too narrow, then even a small amount of noise tends to disperse peaks across a number of bins. On the other hand, if the bin widths are too wide, then the frequency selectivity of the ZCPA is reduced. For a noise-free pure tone, the frequency resolution depends *entirely* upon the resolution of the histogram. For instance, if the tone contributes to a histogram bin that spans 695 Hz – 727 Hz, then it is impossible to tell from the ZCPA which component in this range contributed to the spectrum.

3.3.2 ZCPA with Narrowband Parameters

In order to configure the ZCPA to detect slow-varying tonals at low SNRs, it is quite clear that we must increase the frequency resolution of both the linear filterbank and zero crossing / interval histogram analyses. A modified parameter set is proposed in Table 3.3. The filterbank now covers the 0 kHz–1 kHz band and is uniform, with filter centres spaced 8 Hz apart. Similarly, the histogram is divided into 501 bins, each 2 Hz wide, and consequently, the fine structure analysis is four times denser than the initial filterbank analysis.

PARAMETER TRADE-OFFS IN THE ZCPA

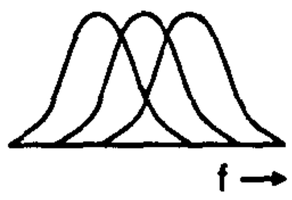
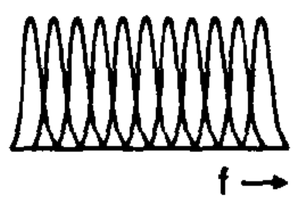
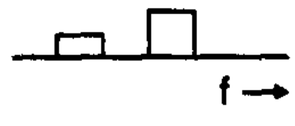
(Fourier) 	Highly-overlapping, wideband filters preserve rapid modulations in signal components, but have very a poor post-analysis SNR and are susceptible to interference between components.
	A bank of narrowband filters maintains a high SNR and can resolve closely-spaced components. However, it may over-resolve modulated signals, has poor time resolution, and is expensive to compute.
(ZCPA) 	A ZCPA compiled from a few recent intervals shows an up-to-date snapshot of a component's frequency. However, in noisy conditions, a few intervals may be unreliable.
	A ZCPA compiled from many intervals reveals with greater clarity which intervals are more frequent. However, if the component frequency is moving, the intervals may now be smeared over many bins.
	Very narrow histogram bins are able to resolve frequencies to arbitrarily fine precision. However, a small amount of jitter in the component frequency causes peaks to disperse.
	Wide histogram bins collect similar intervals together into peaks, at the expense of reducing the final resolution of the ZCPA somewhat.

Table 3.2: An illustrated list describing the beneficial and adverse effects that accompany changes in the key parameters of the ZCPA. (The lower four plots are reciprocal-interval histograms.)

NARROWBAND PARAMETERS	
Parameter	**Value**
channel range	8 Hz – 1000 Hz
channel resolution	125 gammatones, linearly-spaced
bandwidths	uniform, 10 Hz
histogram range	0 Hz – 1000 Hz
histogram resolution	501 bins, linearly-spaced

Table 3.3: Narrowband ZCPA parameters. (The interval window duration, compression function and ZCPA sampling interval are unchanged from Table 3.1.)

Figure 3.13 depicts graphically the difference in resolution between the auditory-like and narrowband parameter set for the 600 Hz–800 Hz section of the filterbank. The filter magnitude responses and histogram bins—plotted as an array of cells above the filterbank response—are considerably narrower in the latter. It should be noted that the final resolution of this ZCPA (i.e., 2 Hz cells) is still lower than that typically employed in a narrowband sonar FFT. Higher-resolution analyses will be discussed shortly.

Figure 3.14B shows the ZCPA time-frequency image and summary pseudospectrum that result from applying the ZCPA with the parameter set given in Table 3.3 to the same four-second signal used to produce Figure 3.12. The tonal structure is now visible, and there is evidence of a harmonic complex with a 50 Hz fundamental. One may compare this with the log-magnitude spectrogram based on a short-time Hann-windowed FFT, shown in Figure 3.14A.

The resolution of the FFT analysis is designed to be roughly commensurate with that of the gammatone filterbank, e.g., the 8 Hz-wide FFT bins match the spacing of the ZCPA filters. Because the magnitude responses of the Hann window and gammatone envelope differ in shape, there are various ways of aligning bandwidths[1]. In the present case, the 3 dB bandwidths of the Hann and gammatone windows are 11.25 Hz and 8.68 Hz, respectively. The selectivity of the ZCPA filterbank therefore exceeds that of the FFT. However, the equivalent noise rectangular bandwidth of the Hann and gammatone windows are 12.00 Hz and 19.63 Hz, respectively. Thus, the post-analysis SNR of the FFT is higher than the ZCPA filterbank.

Comparing Figures 3.14B and 3.14A, the ZCPA appears more sharply-defined than the FFT. This apparent increase in resolution can be attributed to the fact that each FFT filter indiscriminately assigns all energy measured in its output to a single, broad bin, whereas the ZCPA sub-differentiates on the basis of fine structure. Thus, while the FFT stops short of distinguishing between a bin driven by high-energy noise and a bin driven by a tone, the ZCPA goes on to draw this distinction effectively by detecting the ordering influence of a steady signal on the zero crossings of narrowband signals.

[1] See Harris (1978) for a discussion of window functions and their properties.

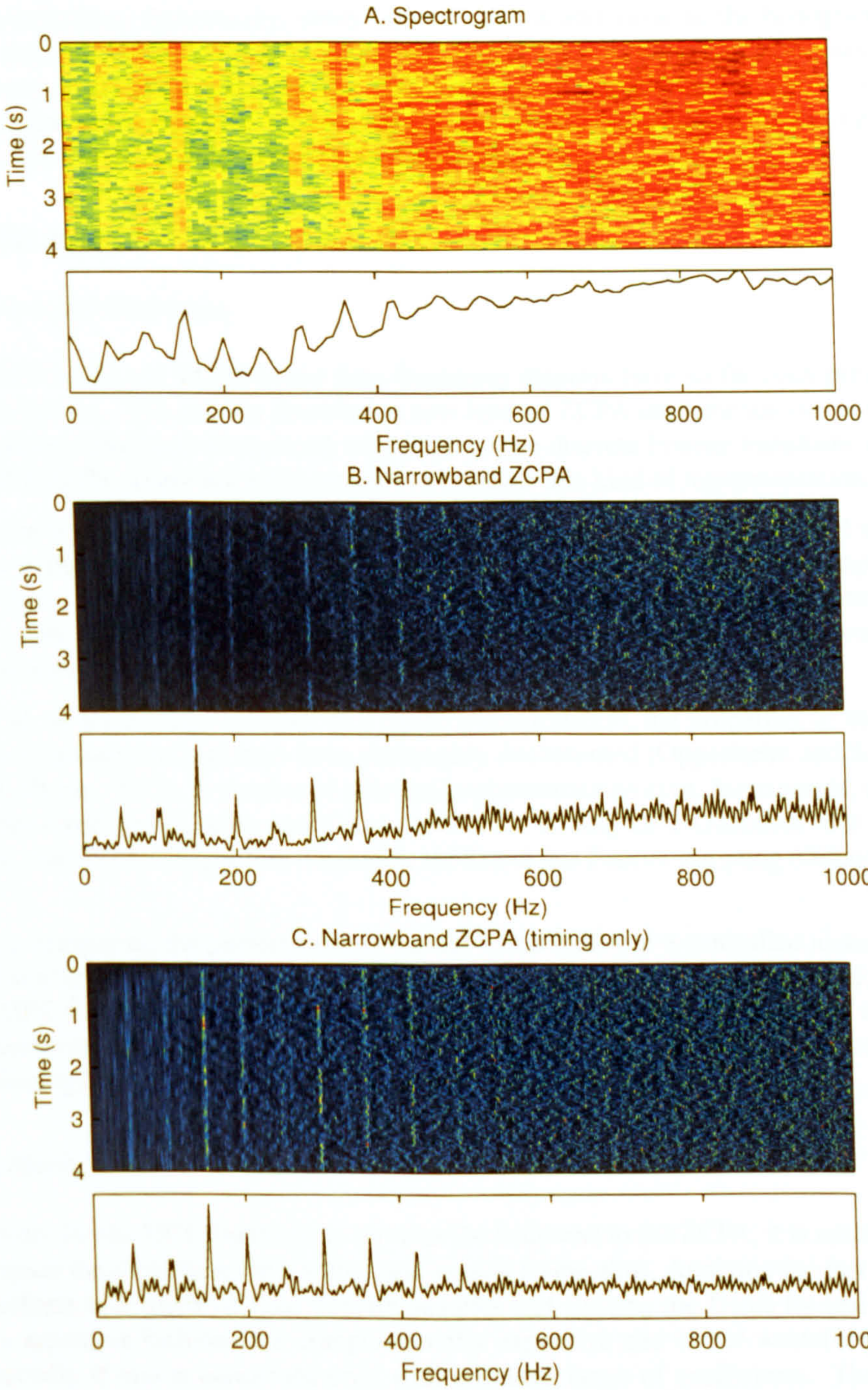


Figure 3.14: Three waterfall time-frequency displays for the first four seconds of the signal shown in Figure 3.4 based on the A) narrowband log-magnitude DFT; B) narrowband ZCPA (log peak compression); and C) narrowband ZCPA (timing only). A summary pseudospectrum is plotted beneath each image.

Lastly, Figure 3.14C shows a narrowband ZCPA generated using a timing-only parameterisation. Specifically, every interval adds a unit value to the histogram bin rather than the compressed peak amplitude, $\log(x + 1)$. It must be emphasised that synchrony capture—inspired by the phase-locking of auditory nerve fibres—is the sole mechanism responsible for revealing narrowband components in the timing-only ZCPA. Quantities relating to power are discarded.

3.3.3 A ZCPA algorithm based on the DFT

Overview and Motivation

The DFT-based and ZCPA-based time-frequency displays have so far been presented as alternatives. This section describes a new kind of ZCPA implementation, in which the auditory filterbank is replaced with a short-time discrete Fourier transform, called the *DFT-ZCPA*. There are a few reasons to consider this kind of implementation.

First, having reconfigured the gammatone filterbank to a uniform distribution of narrow filters in the preceding section, replacing the filterbank with a DFT at this stage only constitutes a rather minor change. In fact, by suitable choice of window function, the DFT could conceivably approximate a uniform gammatone filterbank, with some mild restriction on the filter centres and bandwidths.

Second, as a particular instance of a linear transformation, the properties of the DFT are well understood and have been thoroughly documented (Oppenheim and Schafer, 1989; Harris, 1978). A number of efficient implementations exist, for example, the fast Fourier transform (Cooley and Tukey, 1965), the sliding DFT (Jacobsen and Lyons, 2003), the Goertzel algorithm (Goertzel, 1958) and fast Fourier sampling (Gilbert et al., 2008).

Third, comparing the performance of the DFT and ZCPA in a controlled manner will be more straight-forward, once the ZCPA incorporates the DFT as a front-end. Using the DFT for analysis in the ZCPA will remove any ambiguity concerning whether "commensurate bandwidth" refers to matching SNR (i.e., the rejection of noise) or scalloping loss (i.e., the attenuation of components between filters).

The Sliding DFT

In order for the DFT filterbank to serve as the front-end to the ZCPA, it is necessary to compute the short-time DFT with a one-sample frame-shift. Applying the fast Fourier transform to a block of time domain samples and shifting the frame by one sample in a repetitive fashion is a computationally expensive and rather wasteful process, especially if one is concerned only with a limited range of coefficients. The notion of updating the DFT coefficients using properties of the Fourier transform has already been presented as the *sliding DFT algorithm*, a overview of which is provided by Jacobsen and Lyons (2003). The stages of the sliding DFT are outlined next.

First, a frame of time domain samples with the same length as the DFT, (i.e., N), is buffered, and it is assumed that this buffer runs from time $n-N$ to $n-1$. In addition,

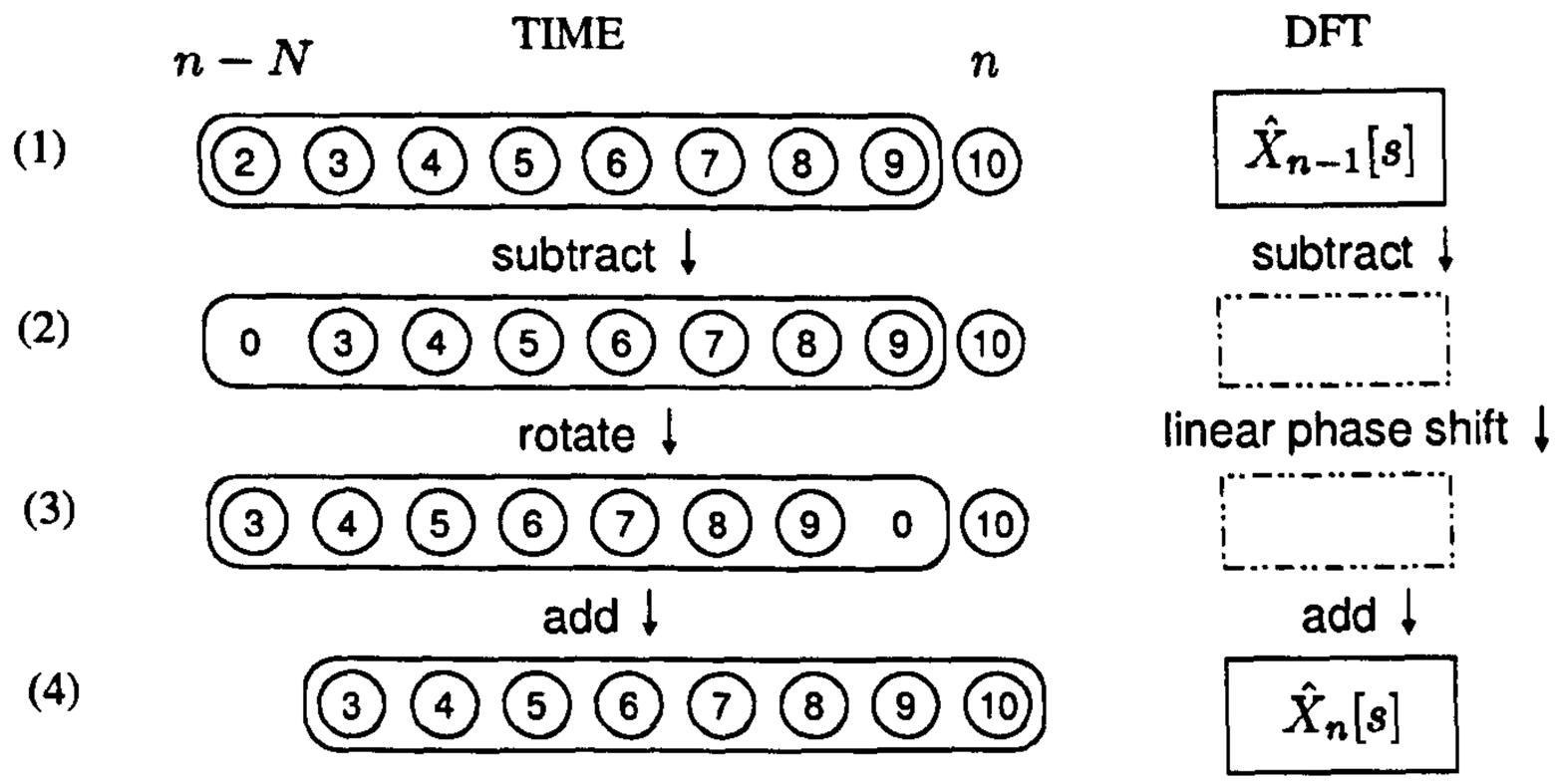


Figure 3.15: One step in the sliding discrete Fourier transform. Operations in the time domain that effectively shift the analysis window along one sample (left-hand side), and the corresponding operations in the frequency domain (right-hand side).

we assume that the N-point DFT for these samples is available. The top row of Figure 3.15 portrays an 8-sample buffer running from $n = 2$ to $n = 9$. The goal is to shift the window along by one sample, and, with as little computational effort as possible, modify the DFT accordingly to match the new buffer content.

The initial step is to set the first sample of the buffer to zero by subtracting a signal that is zero everywhere, except for the first sample, equal to $x[n-N]$. This gives the second row in Figure 3.15. Next, the samples in the buffer are rotated left—that is, one sample backward in time—so that the final sample is now zero, as the third row of Figure 3.15 depicts. Finally, the last sample in the buffer is set to $x[n]$ by addition, producing the final row. Dropping a sample out the left-hand side of the buffer, shifting all the samples in the buffer to the left, and drawing in a new sample from the right, can either be interpreted as sliding the signal under the window, or sliding the window over the signal.

The efficiency of the sliding DFT algorithm is due to the fact that the three elementary time-domain operations invoked above—subtraction, circular shift and addition—have simple counterparts in the frequency domain—namely, subtraction, multiplication by a linear phase shift and addition (Oppenheim and Schafer, 1989). Furthermore, because these operations can be carried out on individual coefficients, if it is only required to track a sub-bank of filters, then only the affected bins need to be updated, as opposed to the entire DFT. Thus, if the s-th complex DFT bin at time $n-1$ is denoted $\hat{X}_{n-1}[s]$, then, advancing one sample, the bin is updated according to the following rule:

$$\begin{aligned} \hat{X}_n[s] &= (\hat{X}_{n-1}[s] - x[n-N]) \exp(i2\pi s/N) + x[n] \exp(i2\pi s/N) \\ &= (\hat{X}_{n-1}[s] - x[n-N] + x[n]) \exp(i2\pi s/N). \end{aligned} \tag{3.43}$$

Conventional DFT implementations, such as the FFT, reduce spectral leakage by multiplying the time domain samples by a tapered window function. The sliding DFT cannot practically accommodate this step, as it is not computed over an entire

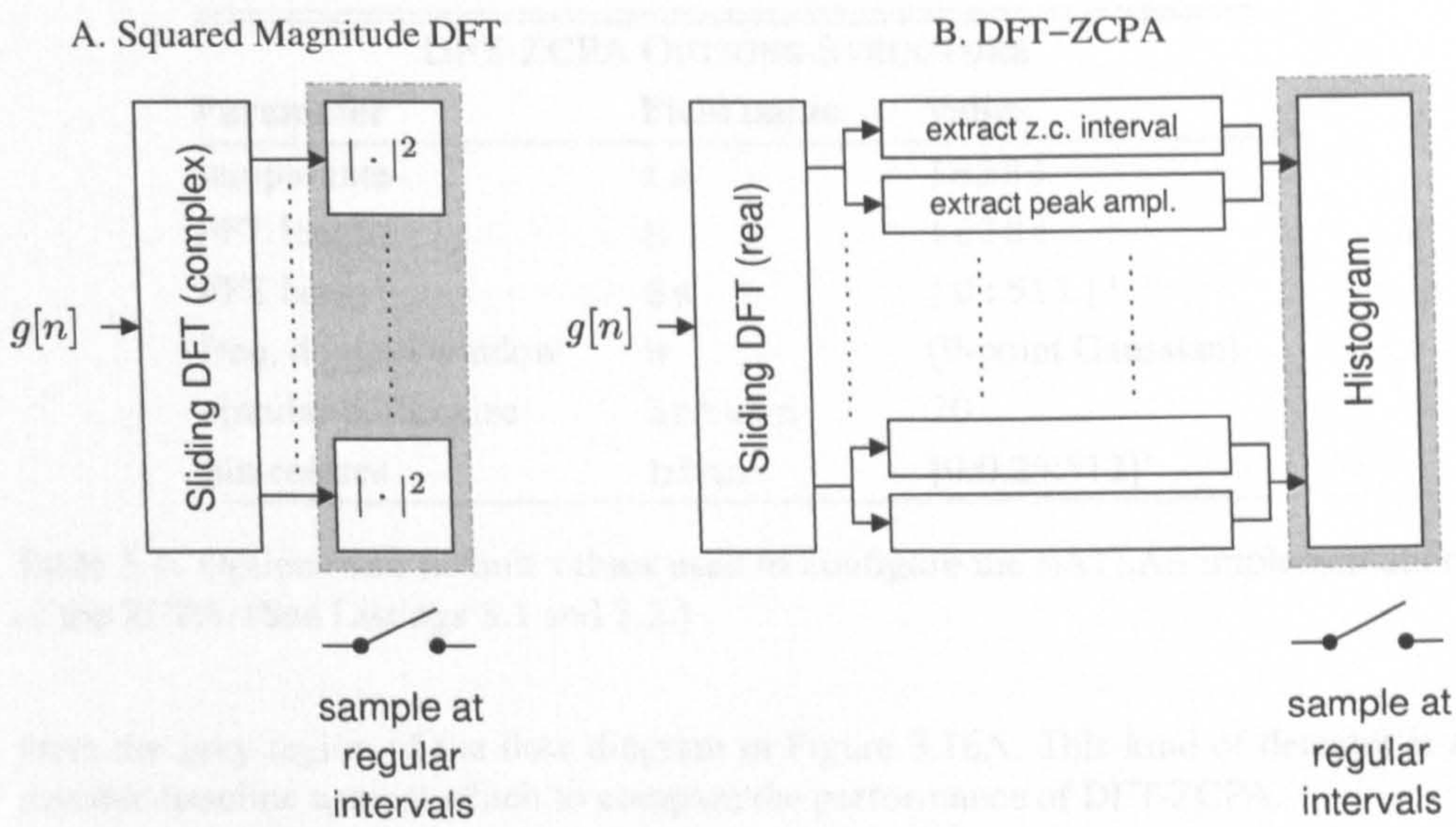


Figure 3.16: A) the squared magnitude DFT and B) the DFT-based ZCPA.

block, but rather sample by sample. Instead, by exploiting the fact that multiplication in the time domain is equivalent to convolution in the frequency domain, windowing is accomplished by convolving an unwindowed DFT with the DFT of the window function.

A few comments about this windowing procedure are in order. First, the coefficients of the window function are generally complex; thus, the frequency-domain convolution entails the addition and multiplication of two complex sequences. (Alternatively, the window function can be carefully chosen to guarantee real DFT coefficients.) Second, because the window function is, in effect, a long, moving-average filter, the magnitudes of the corresponding frequency domain coefficients tend to be significant only over a small region around zero; outside this region, the coefficients can be set to zero. The extent to which the convolution sequence is truncated represents a trade-off: a shorter sequence reduces computational work, but a longer sequence reduces overshoot in the impulse response. Finally, the sliding DFT implementation requires that the convolution be non-destructive, as the unwindowed DFT coefficients must be kept for subsequent iterations.

Processing the Fine Structure

Thus far we have examined (i) how the complex coefficients of the sliding DFT are continually updated as new samples arrive, and (ii) how windowing is effected in the frequency domain. In summary, the sliding DFT effectively provides an efficient, complex-valued, rolling spectrogram. In conventional narrowband sonar processing, the detector operates on the squared magnitude of a DFT sample, or the average squared magnitude of several samples. The test statistic provided to a power detector is drawn

DFT-ZCPA OPTIONS STRUCTURE

Parameter	Field name	Value
sample rate	`fs`	`16384`
FFT length	`N`	`16384`
FFT bins	`Ss`	`[0:511]'`
freq. domain window	`W`	(9-point Gaussian)
circular buffer size	`intwin`	20
bin centres	`bins`	[0:0.25:512]'

Table 3.4: Options and default values used to configure the MATLAB implementation of the ZCPA. (See Listings 3.1 and 3.2.)

from the grey region of the flow diagram in Figure 3.16A. This kind of detector is a suitable baseline against which to compare the performance of DFT-ZCPA.

In the ZCPA, the output of the filterbank undergoes a second stage of processing to extract a combination of temporal and amplitude information from the fine structure. Every time the windowed DFT is computed, each bin is checked to see whether a positive-going zero crossing has occurred in the real part. If so, the interval duration and peak squared amplitude are stored as a pair in a circular buffer associated with that bin. Finally, at regular intervals, the ZCPA histogram is compiled from the reciprocal intervals collected from all the circular buffers, each interval weighted by its peak squared amplitude. Figure 3.16B shows an abstract schematic representation of this algorithm.

Structuring the Algorithm

The DFT-ZCPA implementation discussed in this section comprises four parts: i) a set of options; ii) an internal state; iii) a function which updates the internal state as one sample arrives, and iv) a function which converts an internal state into a ZCPA spectrum, or series of successive ZCPA spectra. The fields of the options structure are set out in Table 3.4. The values in this structure govern the operation of the ZCPA and are not modified at any stage. The fields of the state structure are listed in Table 3.5. Unlike the options structure, the state structure is updated every time a new sample arrives, according to the sliding DFT and fine structure processing described above. This structure completely describes the internal state of the algorithm, so individual invocations of the ZCPA functions can be chained together to process long signals by passing the state structure between calls.

Listing 3.1 describes a cut-down implementation of a MATLAB function, `minzcpa_upd`, which updates the state structure, `st`, following the arrival of a new time-domain sample, `x`. Note that this function implements the sliding DFT and updates the circular buffers, but does not compile the histogram. Instead, a second function, `minzcpa_rec` (Listing 3.2), generates the ZCPA using the contents of a state structure. The MATLAB code in Listings 3.1 and 3.2 is highly inefficient and is

DFT-ZCPA STATE STRUCTURE

Parameter	Field name	Size
sample buffer	`buf`	`N`
complex, unwindowed DFT	`X`	`|Ss|`
real part of windowed DFT	`Xw`	`|Ss|`
coarse interval	`ci`	`|Ss|`
fine interpolation	`fi`	`|Ss|`
recent maximum	`mx`	`|Ss|`
circular buffer (intervals)	`cb.ins[]`	`|Ss|×intwin`
circular buffer (peaks)	`cb.pks[]`	`|Ss|×intwin`

Table 3.5: State variables in the MATLAB implementation of the ZCPA. Every variable listed identifies an array whose size is initialised according to the options structure. Note that the circular buffers are two-dimensional arrays, e.g., each of 512 bins records 20 intervals. The default options set is given in Table 3.4.

written in this compact way in order to convey the basic processes to the reader in an accessible fashion. In fact, the algorithm used to produce all the subsequent ZCPA figures was coded in MATLAB-executable C (MEX) and made proper use of circular buffers, batch processing, appropriate data types and pre-tabulation of trigonometric values. Nevertheless, the C implementation is functionally equivalent to that given in Listings 3.1 and 3.2.

Displaying the DFT-ZCPA for a Synthetic Signal

Before proceeding further, it is appropriate to test the DFT-ZCPA implementation by synthesing a mixture of signal and noise and examining the resulting image. The test signal has been chosen to represent three classes of narrowband signal that are relevant to the analysis of engine tonals: a tone that is resolved between DFT bins, a tone that is weak in relation to the others (its power reduced by 6 dB), and a tone that is phase-modulated. The mixture is synthesised according to the following formula, where $n = 0, \ldots, 196607$ and $f_s = 16384$ Hz:

$$g[n] = \underbrace{\sin(2\pi \cdot 200.4\, n/f_s)}_{\text{unresolved}} + \underbrace{\frac{1}{2}\sin(2\pi \cdot 210\, n/f_s)}_{\text{weak}} + \underbrace{\cos\left(2\pi/f_s \sum_{j=0}^{n} \dot{\phi}[j]\right)}_{\text{modulated}}. \tag{3.44}$$

The instantaneous frequency of the third mixture term is given by

$$\dot{\phi}[n] = 205 + 2\sin(2\pi \cdot 0.25\, n/f_s). \tag{3.45}$$

The function $g[n]$ is added to white Gaussian noise samples with a variance of one hundred. As a guideline, this leads to a signal-to-noise ratio of approximately 16 dB,

```
function st = minzcpa_upd(x,opt,st)
% Update the ZCPA state.
%    MINZCPA_UPD(X,OPT,ST) computes a new state ST for
%    input sample X, options structure OPT and previous
%    state ST.

% Perform sliding DFT
st.X = (st.X - st.buf(1) + x) .* exp(i*2*pi*opt.Ss/opt.N);
st.buf = [st.buf(2:end); x];

% Perform windowing
l  = floor(length(opt.W)/2);
Xn = conv(st.X, opt.W);
Xn = Xn(l+1:end-l);

% Find zero crossings
for n = 1:length(opt.Ss)
    if st.Xw(n) < 0 && real(Xn(n)) >= 0
        % Interpolate interval
        izc = st.Xw(n) / (st.Xw(n) - real(Xn(n)));
        int = st.ci(n) + 1 - st.fi(n) + izc;

        % Store interval in circular buffer
        st.cb(n).ins = [st.cb(n).ins(2:end); opt.fs/int];
        st.cb(n).pks = [st.cb(n).pks(2:end); st.mx(n)];

        % Reset
        st.mx(n) = 0;
        st.ci(n) = 0;
        st.fi(n) = izc;
    else
        % Store maximum; increase coarse interval
        st.mx(n) = max(st.mx(n), real(Xn(n))^2);
        st.ci(n) = st.ci(n) + 1;
    end
end

% Store more recent windowed DFT
st.Xw = real(Xn);
```

Listing 3.1: A MATLAB function that updates the ZCPA state structure, `st`, according to a set of options (`opt`) and an incoming time domain sample (`x`).

```
function zcpa = minzcpa_rec(opt, st)
% Record the ZCPA.
%     MINZCPA_REC(OPT, ST) records the ZCPA corresponding to the
%     internal state ST and the set of options in OPT.

% Reserve space for histogram
zcpa = zeros(size(opt.bins));

% Access left bin centre and bin differences
bl = opt.bins(1);
bd = diff(opt.bins(1:2));

% Compile all intervals into a weighted histogram
for n = 1:length(opt.Ss)
    for m = 1:opt.intwin
        bi = floor((st.cb(n).ins(m) - bl) / bd + 0.5);
        if bi >= 0 && bi < length(zcpa)
            zcpa(bi+1) = zcpa(bi+1) + st.cb(n).pks(m);
        end
    end
end
```

Listing 3.2: A MATLAB function that records a ZCPA pseudospectrum from the contents of the circular buffers in the state structure (`st`). The options structure (`opt`) is also required to specify the histogram configuration.

for a tone with unit amplitude centered on an unwindowed DFT bin. (A more detailed mathematical treatment follows in later chapters.)

The power-based Fourier spectrogram, shown in Figure 3.17A, conforms to the image that one would anticipate, given the description of the signal and the resolution of the DFT. The tone with frequency 200.4 Hz falls between the DFT bins, the centres of which correspond to integer frequencies, so the energy is inevitability resolved across many bins—principally, the 200 Hz and 201 Hz cells. (The window function ensures that the energy is confined to a region around the tone frequency.) The tone centred at 210 Hz is resolved on the centre of a DFT cell but is four times less powerful and, as a result, disappears beneath the noise floor in some rows of the spectrogram. The third signal, a phase-modulated tone whose frequency varies sinusoidally around 205 Hz, is blurred over two or three analysis cells at each time step.

Figure 3.17B shows the DFT-ZCPA image for the signal, in which the underlying analysis resolution is the same as the DFT analysis, and the intervals have been weighted with the peak squared amplitudes. The noise floor of the DFT-ZCPA seems better-suppressed than that of the Fourier spectrogram, in that the background of the colour image appears less mottled. This supports an earlier observation that the zero

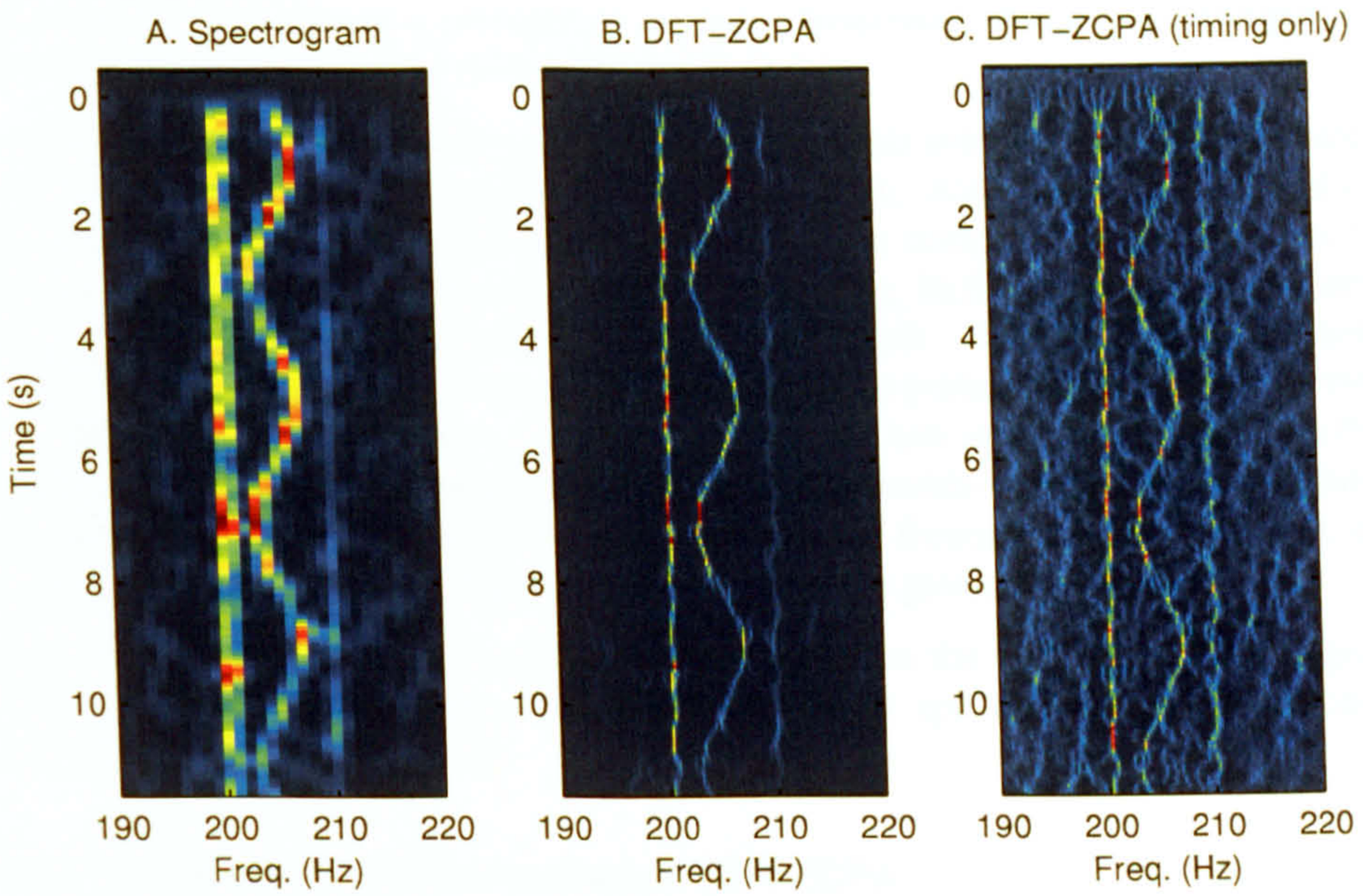


Figure 3.17: Three time-frequency images generated for an additive mixture of the signal described in (3.44) and (3.45), and Gaussian noise. A) squared magnitude of the windowed DFT; B) DFT-ZCPA with squared peak amplitudes; C) DFT-ZCPA based exclusively on timing.

crossing intervals of noise-driven filters are dispersed across many bins, whilst those of steady signals tend to reinforce peaks (*cf.* §2.2.5 and §3.3.1).

A second notable feature of this particular DFT-ZCPA is the relative sharpness with which the narrowband components are resolved, including those whose frequency is non-stationary or falls between DFT bins. Although no such exercise has been undertaken as part of this study, it would be enlightening to compare the accuracy of frequency estimates based on inspection of the two types of image by human operators. A casual glance suggests—to the author, at least—that the delineation of components in the DFT-ZCPA is superior.

The third, qualitative comment is warranted, concerning the effect of additive noise upon the appearance of the components in Figures 3.17A and 3.17B. It is well-known that adding white Gaussian noise to a signal in the time domain is equivalent to adding constant power to its mean squared magnitude Fourier spectrum. At no point is the frequency of a component in the spectrum ever *altered* as a result, although for a given sample function, the peak may be obscured by spurious peaks or reduced in magnitude by destructive interference. However, in the ZCPA, additive noise in the time domain is clearly capable of modifying both frequency (interval) and amplitude (peak) estimates.

The 210 Hz tone demonstrates this point. In the DFT-based spectrogram, although the tonal varies in magnitude and is obscured at times, its frequency is never in doubt, because the signal consistently contributes to the coefficients of the same basis functions, the selection of which is independent of the signal. However, the 210 Hz

tonal in the DFT-ZCPA has a perceptibly unsteady frequency, which is a consequence of the fact that its frequency is derived from noisy data.

The timing-only DFT-ZCPA, shown in Figure 3.17, relies solely on temporal features of the narrowband signals to reveal signal components. Although a background of short artifacts is now apparent in the image, the innate ability of human viewers to process patterns counteracts this disadvantage somewhat. In the terminology of Fulop and Fitz (2006): the signal "detaches" itself from the "froth". Remarkably, the weakest component in the mixture, centred at 210 Hz, appears as prominently as the other two do. In light of the discussion above, it seems likely that a low local SNR destabilises the fine structure of a weaker component and increases its breadth—and hence its visibility, whereas a high local SNR exerts little influence over the frequency of a component, so that stronger, more reliable components are displayed in greater detail.

Having verified that the ZCPA implementation based on the DFT maps a test signal to an appropriate time-frequency image, we shall now apply it to a recorded sonar signal—with some modifications.

3.3.4 Sonar Signals in the Multi-resolution DFT-ZCPA

The demand for a uniform, narrowband analysis at low frequencies, combined with the appeal of the discrete Fourier transform, led to the abandonment of auditory-motivated filter centres and bandwidths, almost as a matter of practical necessity. Nevertheless, for a sonar processor, progressively wider analysis bands may still be favourable at higher frequencies, for at least two reasons. First, the bandwidth of a tonal is usually proportional to its frequency (Burdic, 1984), and Doppler shifts are more pronounced at higher frequencies. In both cases, wider analysis bandwidths are required at higher frequencies to ensure that modulated harmonics are not over-resolved. Second, although the detection of narrowband emissions has been emphasised thus far, a comprehensive sonar display should enable the visualisation of transient events and rapidly-changing high-frequency components, such as the vocalisations of marine mammals. Following these considerations, the *multi-resolution DFT-ZCPA* is now proposed.

Description

The multi-resolution DFT-ZCPA is simply the concatenated output of several simple DFT-ZCPAs blocks, each with its own time-frequency configuration. The edges of the DFT and histogram bins of each block are chosen to align at the edges so that the resulting ZCPA histogram provides coverage of the full band. Many configurations are possible, but in view of the concerns mentioned above, the following guidelines are advisable: i) the DFT cells should grow wider with each block (i.e., the DFT length must be shortened), ii) the frequency with which the ZCPA is sampled should be increased by the same ratio to reflect the increased time resolution, and iii) the ratio of histogram bins should be held constant. Figure 3.18 schematically depicts both flow diagrams for the ZCPA blocks (in abstract) and a time-frequency tiling of the resulting image.

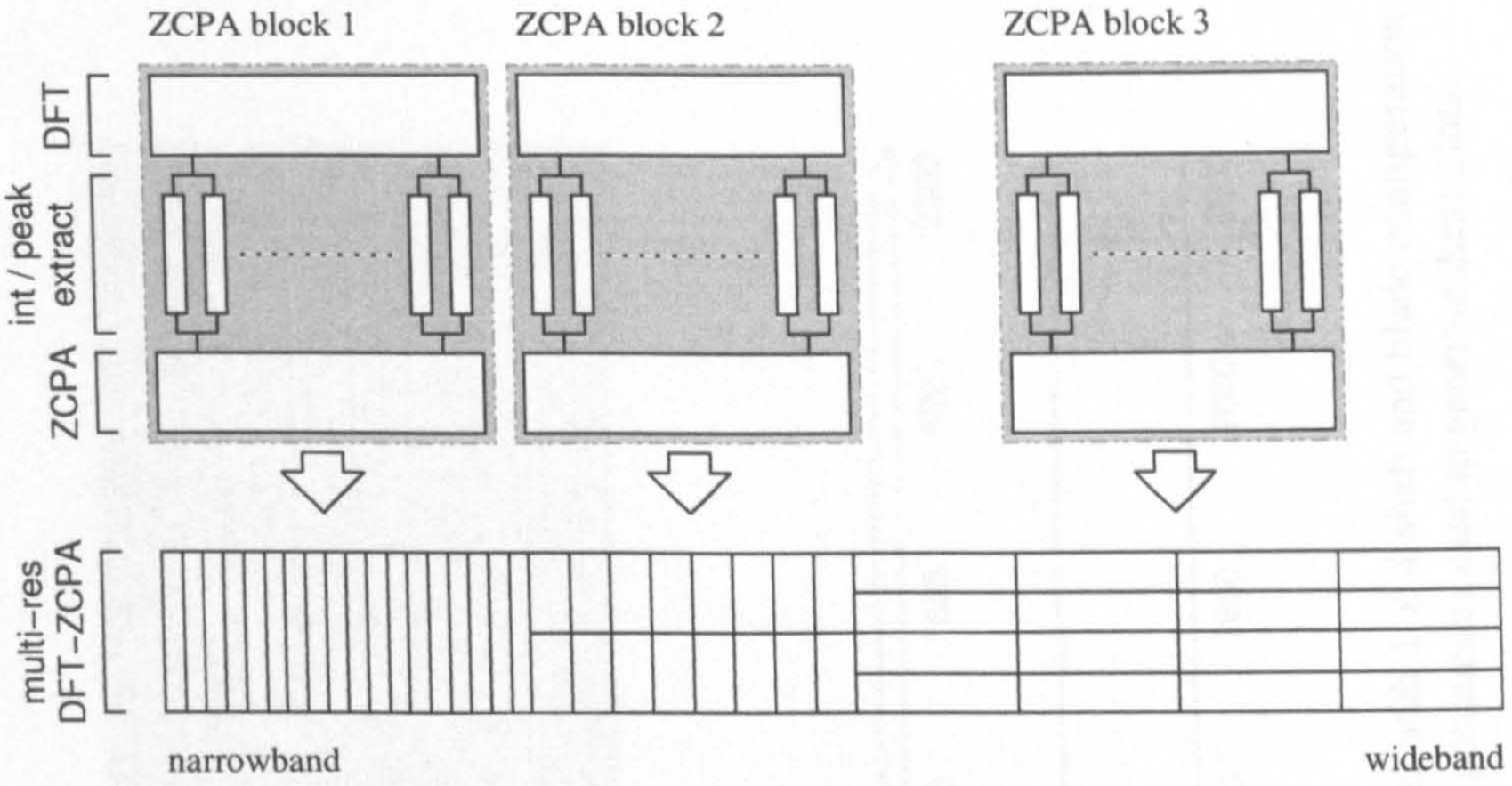


Figure 3.18: The multi-resolution ZCPA. The internal configuration of each ZCPA block (light grey) is designed to recall a 90^o clockwise rotation of Figure 3.16B.

A final issue regarding implementation relates to the rescaling of DFT coefficients to ensure an even response across the DFT-ZCPA (if desired). Because the DFT length varies from block to block, the amount of power delivered to the peak amplitude extraction block varies. The DFT output can be adjusted to ensure either: i) a flat tonal response, i.e., two tonals of equal power register equal height in the ZCPA; or ii) a flat white noise response, i.e., the time-averaged ZCPA for white noise input is roughly constant and does not contain jumps at the block boundaries. As sonar signals contain a large broadband component, the most natural choice is (ii).

Displaying the Multi-resolution DFT-ZCPA for a Recorded Sonar Signal

Figure 3.19 demonstrates the multi-resolution DFT-ZCPA for twelve seconds of a sonar recording, along with the summary ZCPA and a summary log-magnitude spectrogram. The parameter set used to produce the image is summarised in Table 3.6 and adheres to the general criteria set out above. The narrowband peaks are compressed using the $\log|x+1|$ function.

The tonals are visible in the DFT-ZCPA image as vertical lines, the faintness of which appears to result from their width, rather than their intensity. In the summary plot, the tonals stand out against the noise floor and their frequencies are located with greater precision than in the summary DFT spectrogram shown beneath. (The DFT spectrogram image has been omitted.)

3.3.5 Statistical Performance Analysis of the DFT-ZCPA

Throughout this chapter, qualitative, visual assessments such as, "The tonals appear ill-defined in the DFT," or, "The discrete lines are easier to spot against the noise

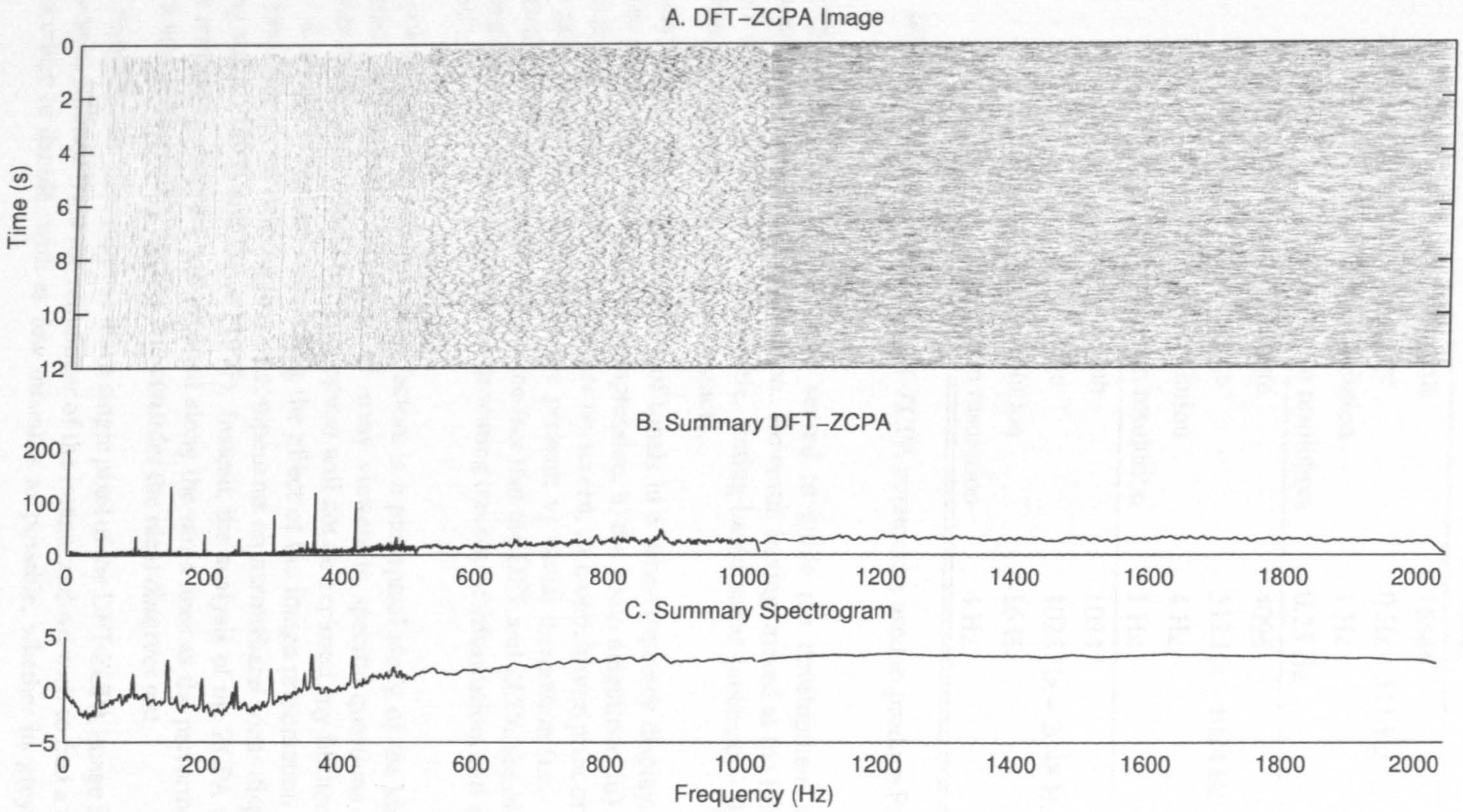


Figure 3.19: Multi-resolution DFT-ZCPA for single-channel sonar recording of a large oil tanker with 1×4 shaft and blade configuration propelling at 100 rpm. A spectrogram for this vessel is shown in Figure 3.4. See Table 3.6 for the parameters used to generate these plots.

	MULTI-RESOLUTION DFT-ZCPA	
Block	**Parameter**	**Value**
1.	DFT length	16384
	DFT range	0 Hz – 511 Hz
	DFT resolution	1 Hz
	Histogram resolution	0.25 Hz
2.	DFT length	4096
	DFT range	512 Hz – 1024 Hz
	DFT resolution	4 Hz
	Histogram resolution	1 Hz
3.	DFT length	1024
	DFT range	1025 Hz – 2048 Hz
	DFT resolution	16 Hz
	Histogram resolution	4 Hz

Table 3.6: Multi-resolution DFT-ZCPA parameters used to produce Figure 3.19.

background in the ZCPA," have served to guide the development of the ZCPA towards a realistic sonar application. However, having arrived at the multi-resolution DFT-ZCPA—a tentative, admissible, timing-based sonar processor—these kinds of subjective remark now form an obstacle.

Many factors affect the visibility of tonals in a time-frequency display, for instance, i) the colour map and dynamic compression, ii) the zoom magnitude, iii) the means by which the image is viewed, whether on screen, in monochrome print or colour print; iv) prior knowledge that tonals are present; v) visual integration (i.e., the ability to spot lines and other patterns); vi) the fact that the DFT and ZCPA are often presented alongside each other, allowing the unwitting transfer of information; and vii) researcher bias[1].

A possible remedy against these factors is a perceptual study of the kind alluded to earlier, in which the responses of many viewers to specific questions are collected under controlled conditions. This option will not be explored any further in this work. For a review of studies concerning the effect of the image presentation and the state of the observer on detectability of components on narrowband sonar displays, consult Grigorakis (1997) and Dawe (1997). Instead, the analysis of the ZCPA carried out in the remaining chapters will proceed along the same lines as the performance analysis of a narrowband DFT display and consider the ideal observer test.

To illustrate this aim, suppose that a single pixel of the DFT-ZCPA image in Figure 3.19 has been isolated from the remainder of the image, and we wish to find a mathematical procedure to decide, with as few mistakes as possible, whether its greyscale is most

[1] As Francis Bacon observed, "What a man had rather were true he more readily believes."

likely the result of a signal mixed with some noise, or just the noise background on its own. To achieve this, we require minimally: i) a statistical description of the pressure wave at the hydrophone for each hypothesis, and the prior probability of the hypothesis; ii) an understanding of the process by which the greyscale of the ZCPA cell in question is derived from the input signal; and finally, iii) a means to infer from the observed greyscale which hypothesis is most likely.

The establishment of a basic framework to choose between signal hypotheses on the basis of ZCPA measurements opens up further possible applications, including multiple hypothesis testing (e.g., "Is there a 200 Hz signal, a 201 Hz signal or no signal?"), multiple-sample hypothesis testing (e.g., "Given the colour of these five ZCPA cells, should I say a signal is present?"), and, ultimately the estimation and tracking of signal parameters.

Finally, as the ideal observer test is unbiased, it allows us to compare the performance of the DFT and ZCPA in a way that is not contaminated by any of the human factors mentioned earlier. The goal of the next two chapters is to investigate whether this can be done.

3.4 Summary

In an ocean model governed by the linear wave equation, two small forces act on a small volume of seawater at any given instant: inertial forces (due to elemental mass) and elastic forces (due to elemental deformation). A vibrating source communicates its motion to the surrounding water and causes longitudinal waves to radiate outwardly in planar, cylindrical or spherical wavefronts. The geometry of the propagation depends on the directivity of the source, the presence of reflective boundaries such as the sea floor and sea surface, and refraction effects.

A passive sonar receiver stationed at a moderate distance from a target detects a portion of its radiated acoustic energy and uses the distribution of energy in frequency to classify the target. The spectral lines, or *tonals*, generated by rotating and oscillating engine components are particularly salient classification features. The overall acoustic signature must be received against a background of noise, arising from surface waves, rain, remote shipping, industry, sea life and self noise.

Power-based sonar detection applies a threshold to the power measured at the receiver. A processing chain of spatial and frequency-dependent filters removes all the unwanted background noise energy, then statistically independent samples of the squared envelope are averaged together to obtain Gaussian statistics. Using the probability distribution for the power received under noise-only conditions, the probability of false alarm can be determined, then the probability of detection for a particular signal is modelled using the broadband and narrowband passive sonar equations.

In temporal processing, both frequency and power are measured from the signal. This chapter has described the development of the *discrete Fourier transform with zero crossings and peak amplitudes* (DFT-ZCPA), as a possible technique for representing narrowband sonar signals. The DFT block provides a bank of narrow filters, and the ZCPA block uses the fine structure in the zero crossing intervals to generate a sharp spectrum. The question of optimal detection in this representation remains open. Rather than tackling the issue of signal detection in DFT-ZCPA algorithm wholesale, the next chapter examines the principle of detection using one atomic unit of temporal information: the zero crossing interval.

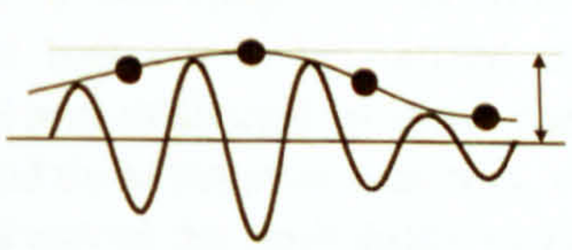
Chapter

4

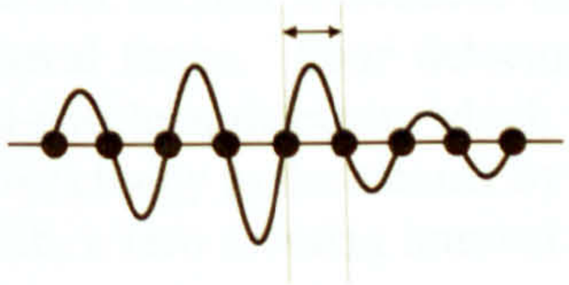
Elementary Interval Detectors

The preceding chapter presented, in broad terms, the idea of using the zero crossing intervals in the output of an auditory filterbank as a means of detecting and tracking tonal components. The processing in a single pathway of the zero crossings with peak amplitudes algorithm includes a series of linear and non-linear transformations, such as zero crossing detection, log compression and histogram formation, making attempts to characterise optimal detection in the ZCPA rather ambitious. As an intermediate step, this chapter pursues the more modest goal of developing a simple class of interval detector and investigating its operation under highly idealised conditions, with a view to elaborating upon the basic model in later chapters.

The detectors described in this chapter must choose exactly one of two hypotheses: under the first hypothesis, H_0, the input to the receiver is due to noise alone; and under the second hypothesis, H_1, the input to the receiver is a mixture of signal and noise. The prior probabilities assigned to H_0 and H_1 are equal, and the receiver is a linear system with a known impulse response. The squared-envelope detector records a single observation of the envelope at the output of the receiver and must choose between H_0 and H_1 in a way that minimises the probability of an incorrect decision. The interval detectors attempt the same classification using the time interval between two successive zero crossings as a test statistic. The quantities that make up the test statistic for the (squared) envelope detector and interval detectors are indicated on the diagrams below.



envelope statistic



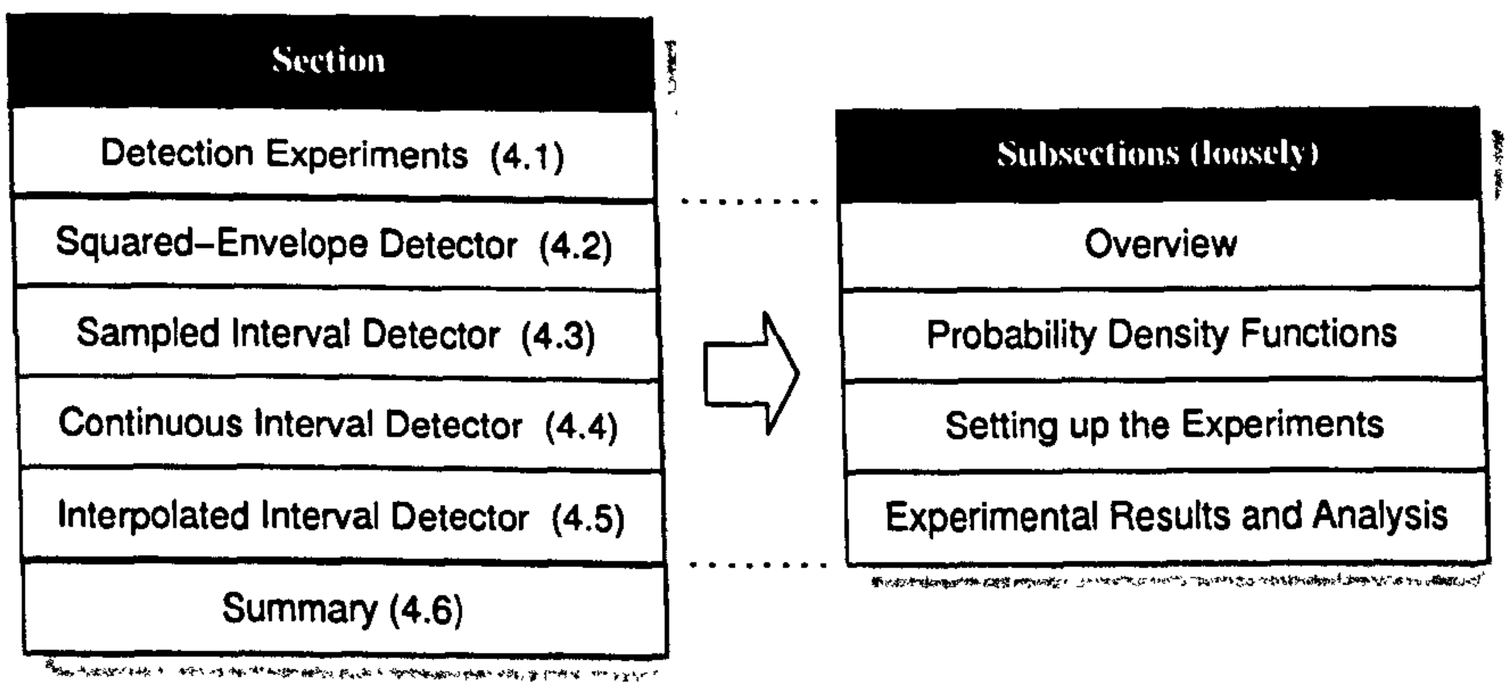
interval statistic

Chapter Scope and Outline

The attention of this chapter is confined in at least five regards, which we shall make clear at the outset. First, the zero crossing with peak amplitudes algorithm forms a histogram using the output of many filters; here, the focus is restricted to the output of a single analysis filter. Second, the ZCPA constructs a histogram from many zero crossing intervals of a filter; the interval detectors in this chapter can only operate on a single interval. Third, the ZCPA weights the contribution of an interval to the histogram using a local estimate of the envelope, whereas the interval detectors presented next operate on timing information alone. Fourth, both the noise-only and signal-and-noise hypotheses must be governed by stationary Gaussian random processes with known parameters. Fifth, while it is customary to evaluate the performance of a detector with respect to a fixed probability of false alarm rate, e.g., using ROC curves and transition curves, the probability of error is the sole performance metric adopted in this chapter.

This list of requirements may at first appear to limit the utility of interval detection; however, this chapter presents an opportunity to rigorously evaluate prototype detectors in noise conditions that are well-understood, before advancing their development any further. (Chapter 5 will examine the possibility of relaxing some of these constraints.) In addition, there are a couple of conventions in this chapter that may potentially be mistaken for limitations of the detection routines themselves. First, the assumption of a white noise background is merely expedient; the detectors may be readily extended to accommodate coloured noise backgrounds on the basis of the material presented in this chapter alone. Second, the assumption of binary detection is particular to our concern, but the detectors can be configured to select one of several hypotheses and even, in that capacity, employed as estimators. (See Section 6.3.1 below.)

Chapter 4 Outline



The outline of this chapter is as follows. The first section introduces the detection experiments that are to be carried out in general terms. Four detectors are then constructed and evaluated in turn: the squared-envelope detector, which serves as a baseline; and three interval detectors, which differ chiefly in the manner by which they extract, and model the probability associated with, a zero crossing interval.

4.1 Detection Experiments

4.1.1 Analysis Filter and Noise Process

The detection tasks reported in this chapter adhere to the same basic format. Each assumes that the signal (if present) and background noise form an additive mixture, which has been received via a linear *analysis filter* prior to detection. The impulse response of the analysis filter is a Gaussian-windowed sinusoid,

$$h_a[n] = \exp\left[-2\left(\frac{\alpha_a n/f_s}{T_a}\right)^2\right]\cos(2\pi f_a n/f_s), \tag{4.1}$$

parameterised by f_s, T_a, α_a and f_a. The first parameter, f_s, is the sample rate, which is a constant 16384 Hz and may be assumed to represent this value wherever it appears in this chapter[1]. T_a controls the overall length of the impulse response; specifically, the window is near zero for $|n| > f_s T_a$. The tapering of the window is tuned separately by the α_a parameter, a suitable choice being $\alpha_a = 2.5$. Finally, f_a is the frequency of the fine structure, in Hertz. From the frequency domain perspective, T_a controls the bandwidth of the filter, α_a regulates the side lobes and f_a selects a centre frequency. The squared-magnitude response of the filter is denoted $|\mathscr{H}_a[s]|^2$. Figure 4.1A provides an example of an impulse response generated using (4.1); Figure 4.1B plots the corresponding squared-magnitude response.

We assume in the first instance that the noise background against which the signal must be detected is white. If the noise variance (or, equivalently, total power) is denoted σ_n^2, the power spectral density of the noise-only random process is given by the product of the noise p.s.d. with the squared-magnitude response of the analysis filter, i.e.,

$$\mathscr{S}_0[s] = |\mathscr{H}_a[s]|^2\sigma_n^2. \tag{4.2}$$

The noise power in a 1 Hz band at both positive and negative frequencies, designated N_0, is found by dividing the total noise power by the baseband width and multiplying by two, i.e., $N_0 = 2\sigma_n^2/f_s$. A flow diagram depicting the synthesis of the noise process is shown in Figure 4.2A. The random processes at the input and output of the analysis filter are denoted G and X, respectively.

4.1.2 Signal Process

The target signal is a random process formed by convolving white Gaussian noise with another Gaussian-windowed sinusoid, associated with what we hereafter refer to as the *signal filter*,

$$h_s[n] = A_s \exp\left[-2\left(\frac{\alpha_s n/f_s}{T_s}\right)^2\right]\cos(2\pi f_c n/f_s). \tag{4.3}$$

[1] Most of the recorded signals provided by QinetiQ were sampled at a rate of 16384 Hz, and, as a power of two, it was a convenient value to continue using.

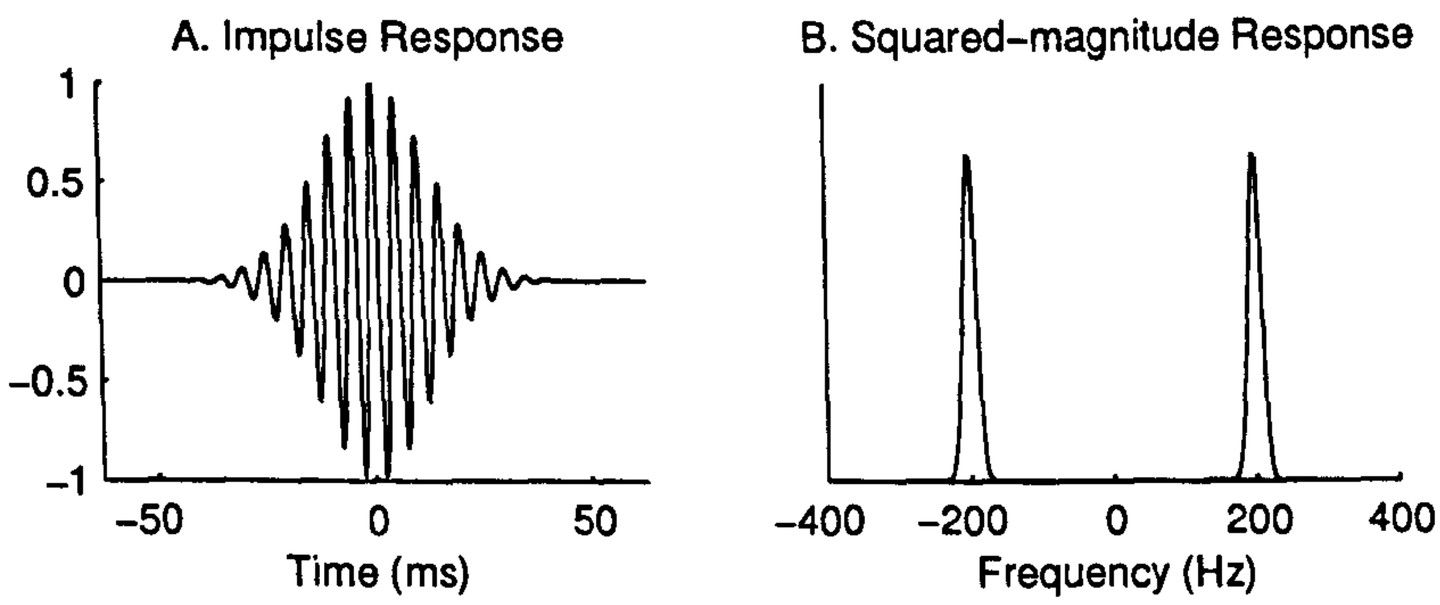


Figure 4.1: A) impulse response and B) squared-magnitude response of the analysis filter described by (4.1) with parameters $f_a = 200$, $T_a = \frac{1}{16}$ and $\alpha_a = 2.5$.

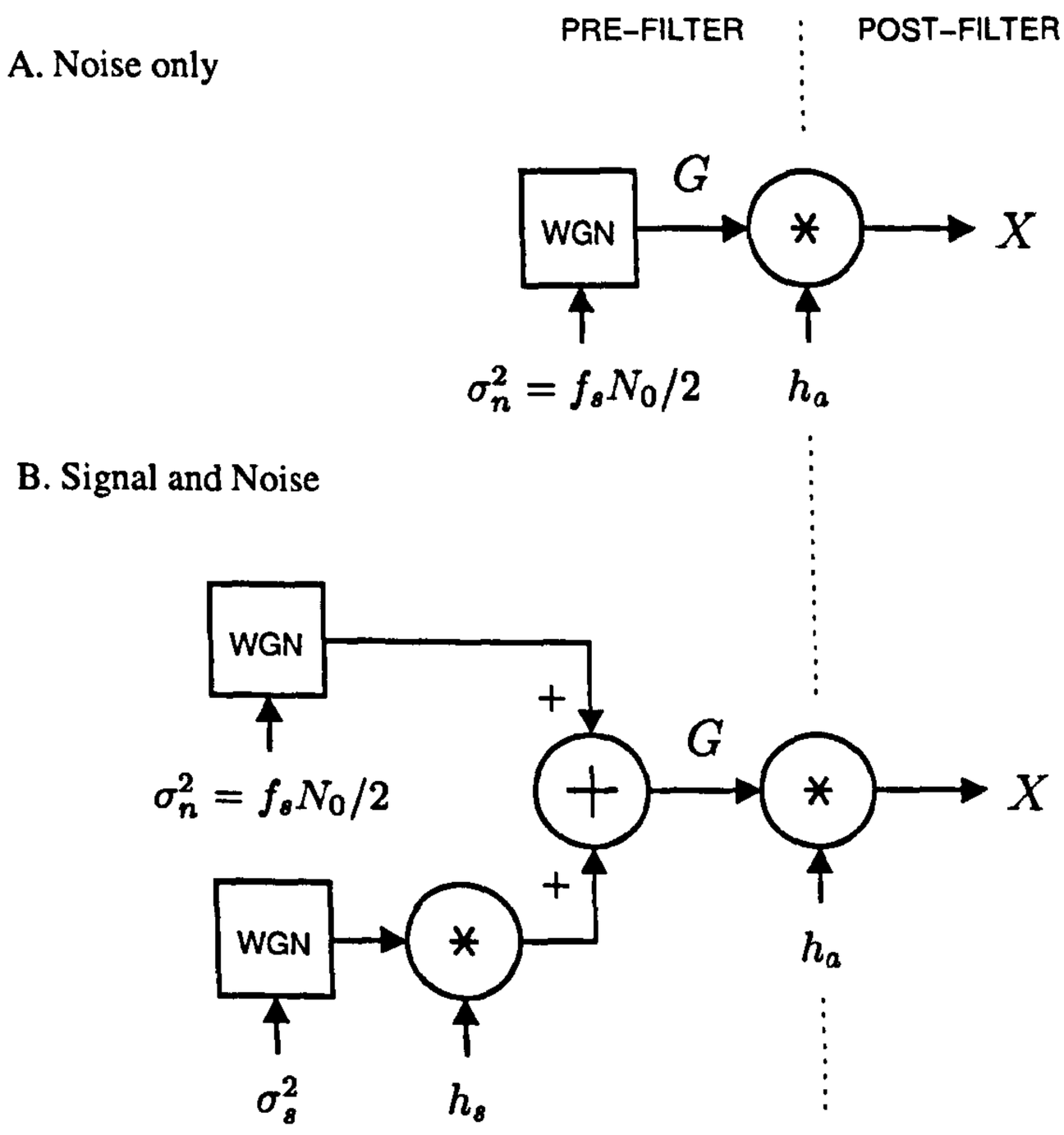


Figure 4.2: Demonstrates how the signals for H_0 and H_1 are generated. The WGN block generates white Gaussian noise with variance controlled by its lower input. The output in each case is a random process denoted G, which forms the input to the detector block diagrams shown in subsequent figures.

(A random process generated by convolving white Gaussian samples with an impulse response is referred to by time series analysts as an *auto-regressive moving average* model.) As before, T_s and α_s respectively control the duration and sharpness of the impulse response. The subscripts on T and α clarify to which signal model they belong ('a' and 's' denote 'analysis' and 'signal', respectively) with one exception: f_s cannot refer to the signal frequency because it is reserved for the sample rate; instead f_c is adopted, with 'c' denoting 'centre'. The constant A_s normalises the impulse response so that

$$\sum_{\forall n} h_s^2[n] = 1. \tag{4.4}$$

This simplifying measure ensures that the total power of the signal process is equal to the variance of the white noise process convolved with the signal filter, σ_s^2. A flow diagram illustrating the stages undertaken to synthesise a signal-and-noise mixture is provided in Figure 4.2B. A narrow notch of noise will serve as a sinusoid-like process that satisfies stationary Gaussian assumptions, until the derivation of the interval distribution for a pure sinusoid is tackled in Chapter 5.

4.1.3 Signal-to-Noise Ratio

The *signal-to-noise ratio* (SNR) measures the relative contribution of signal and noise to an additive mixture in terms of their power. Various definitions of signal-to-noise ratio exist, but the usage is typically applied in one of two senses, depending on the application. In the first usage, signal-to-noise ratio is a quantity that measures how difficult it is to detect a signal in a specific noise background. This is generally the case in automatic speech recognition studies, where we are interested to inquire of a system's performance given a particular SNR. In the second usage, signal-to-noise ratio refers to a system's ability to reject noise and therefore indicates the quality of the receiver. In engineering a radio receiver, for example, one might speak of "adjusting the parameters of the receiver in order to maximise SNR". It is helpful to define explicitly what is meant by SNR in this chapter before moving on.

Global SNR

A *pre-analysis SNR* measures the relative contribution of signal and noise to the random process G. (See Figure 4.2.) One possible choice of pre-analysis SNR is the *global signal-to-noise ratio*, which is defined here as the decibel ratio of total signal power to total noise power in a mixture prior to analysis, i.e.,

$$\mathrm{SNR}_g = 10 \log_{10} \frac{\sigma_s^2}{\sigma_n^2}, \text{ dB}. \tag{4.5}$$

The global SNR is an appropriate expression of the signal-to-noise ratio for broadband detection and speech recognition studies. This definition also identifies the *power* of a sampled signal with the *variance* of its samples, and hence avoids several potential conflicts in language and mathematical notation.

Narrowband SNR

The global signal-to-noise ratio is not a suitable quantity for measuring how difficult it is to detect a narrowband signal, as it provides no insight into the true signal and noise conditions, when removed from the context of a particular sample rate or baseband width. For example, doubling the sample rate causes a 3 dB drop in the global SNR, although the difficulty in detecting the signal is unchanged. Accordingly, the definition of signal-to-noise ratio adopted in this chapter measures the ratio of the total signal power to the noise power in a 1 Hz bandwidth:

$$\begin{aligned} \text{SNR} &= 10 \log_{10} \frac{\sigma_s^2}{N_0}, \text{ dB} \\ &(\approx \text{SNR}_g + 39.13 \text{ dB, if } f_s = 16384.) \end{aligned} \tag{4.6}$$

The quantity N_0 is defined in Section 4.1.1 above. The narrowband SNR is completely invariant with respect to both the choice of sample rate and the analysis filter. Wherever the term SNR appears without any qualification, the pre-analysis, narrowband usage is intended.

Post-analysis SNR

The *post-analysis SNR* refers to the signal-to-noise ratio in the random process X, and it accounts for the effect of the analysis filter upon the mixture. The post-analysis SNR is frequently defined as either i) the ratio of resolved signal power to resolved noise power (Peebles, 1993), or alternatively, ii) the ratio of resolved signal power to resolved noise power in a 1 Hz band (Dawe, 1997). The post-analysis signal-to-noise ratio is conventionally employed in the sonar literature. A pre-analysis (narrowband) SNR is the most suitable for this study, however, as this study investigates methods of detection that are not based on power. Nevertheless, it will occasionally be appropriate to refer to the post-analysis SNR when commenting on experimental results.

4.1.4 Experimental Procedure

The detectors described in the remainder of this chapter are to be evaluated in the same task: the detection of a narrowband signal process against a white noise background. In each experiment, two signals are synthesised and form the input to a detector. The first signal is a sample function of the noise-only process, so each decision of the detector corresponds to either a true negative or false positive. The second signal is a sample function of the signal-and-noise process, and the detector generates either true positives or false negatives. The empirical probability of error is calculated from the true positive, true negative, false positive and false negative counts, and recorded against the set of experimental parameters.

The aim of these experiments is to establish the effect of the independent variables—signal-to-noise ratio, signal frequency and band frequency—upon the probability of error. When a parameter is held constant, it is assigned the value in Table 4.1. To maintain a uniform prior distribution, an equal number of test statistics is extracted for

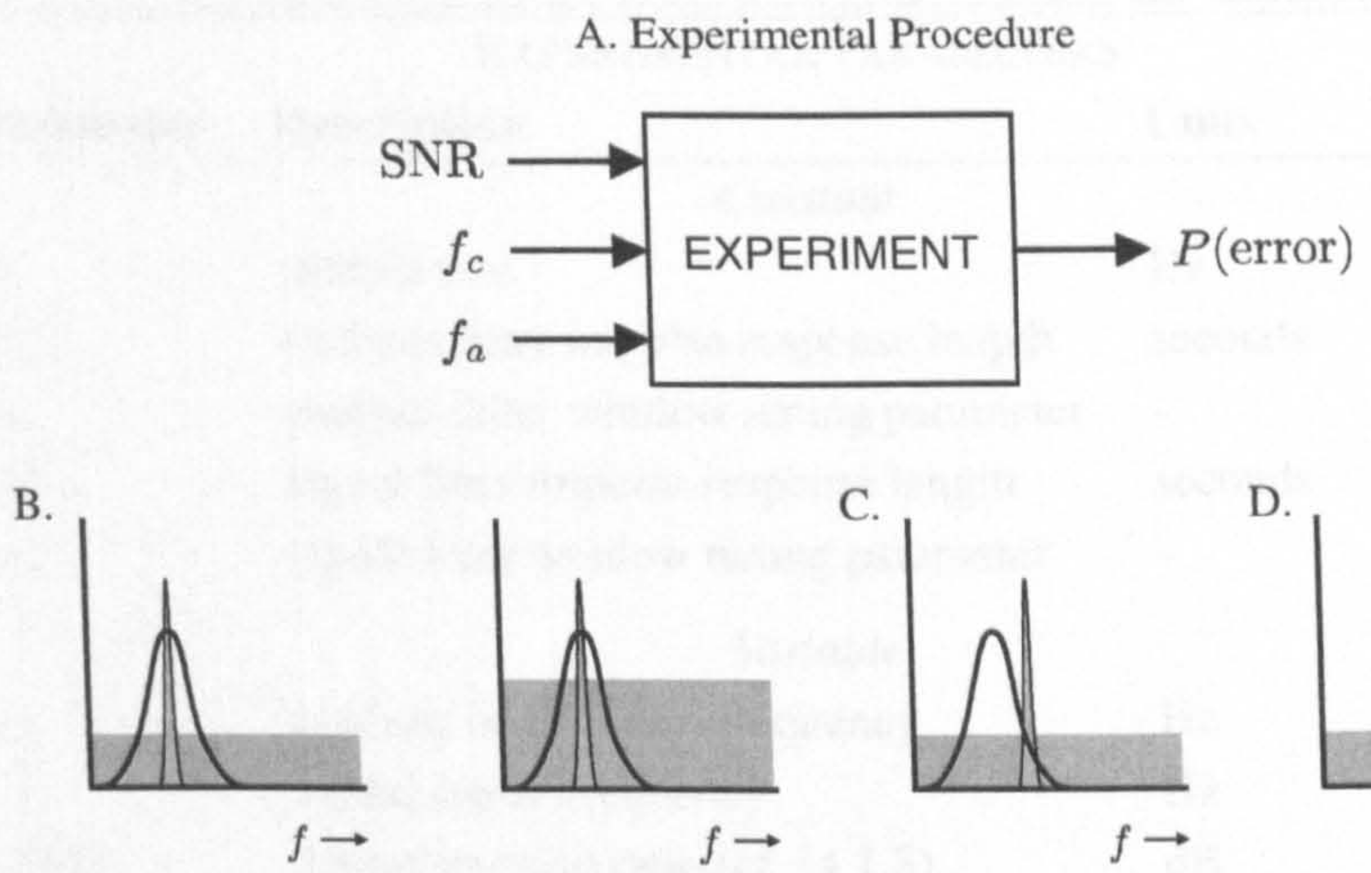


Figure 4.3: A) independent and dependent variables of the experiment; B) detection of a signal at a high and low SNR (left and right); C) detection of a signal moved away from the band centre; D) detection at high frequencies.

H_0 and H_1, and every detector is memoryless, so its performance does not depend upon the order in which the measurements are supplied.

4.1.5 Research Questions

How does the detector's performance vary with SNR? It is necessary to note how the detector performs when the noise level is changed. This will determine whether the detector is better-suited to low or high signal-to-noise ratio applications. In these experiments, f_a and f_c are fixed at the same value, 200 Hz, and the SNR is varied between 0 and 40 dB. This is shown pictorially in Figure 4.3B.

What is the effect of displacing the signal from the band centre? This question is motivated by the apparent importance of 'synchrony capture' in auditory models. It will be addressed: i) by repeating the SNR experiment described above, with the signal displaced from the band centre by 10 Hz and 20 Hz, and ii) by holding the SNR fixed at 20 dB and varying the signal frequency across the analysis bandwidth. See Figure 4.3C.

Does the detection performance relate to the absolute frequency of the signal? The relative error introduced into envelope measurements by round-off error in the CPU is likely to be several orders of magnitude smaller than the relative error introduced into interval measurements by time-domain sampling. The effect of changes in the absolute signal frequency upon error will be investigated by repeating the on-centre detection task for a variety of signal and band centre frequencies across the range 200–1000 Hz. This idea is conveyed in Figure 4.3D.

Is it possible to predict the performance of the detector? There are at least two clear reasons to pursue this question. First of all, if the analytical results agree with those obtained by random trials, it validates the theoretical foundations on which the detector

EXPERIMENTAL PARAMETERS

Parameter	Description	Units	Value
	Constant		
f_s	sample rate	Hz	16384
$2T_a$	analysis filter impulse response length	seconds	$T_a = 1/16$
α_a	analysis filter window tuning parameter	-	2.5
$2T_s$	signal filter impulse response length	seconds	$T_s = 1$
α_s	signal filter window tuning parameter	-	2.5
	Variable		
f_a	analysis band centre frequency	Hz	200
f_c	signal centre frequency	Hz	200
SNR	signal-to-noise ratio (*cf.* §4.1.3)	dB	20

Table 4.1: A list of experimental parameters.

was constructed; or, to state the reverse, if the detector performs better or worse than the theoretical work predicts, then errors in either the theory or the experimental procedure will be exposed at an early stage. Second, if it proves possible to predict the behaviour of the detector reliably, then additional, fine-grained results can be obtained without having recourse to random trials, which yield only approximate results and take a long time to complete.

How do the results of the different detectors compare? One of the ultimate goals of this chapter is to compare the performance of interval-based detectors with that of a conventional power-based detector, with a view to improving on the latter. It is therefore necessary, in addressing this question, to identify the conditions under which the detector fails, to trace the cause of the failure and then, if possible, to design a better detector.

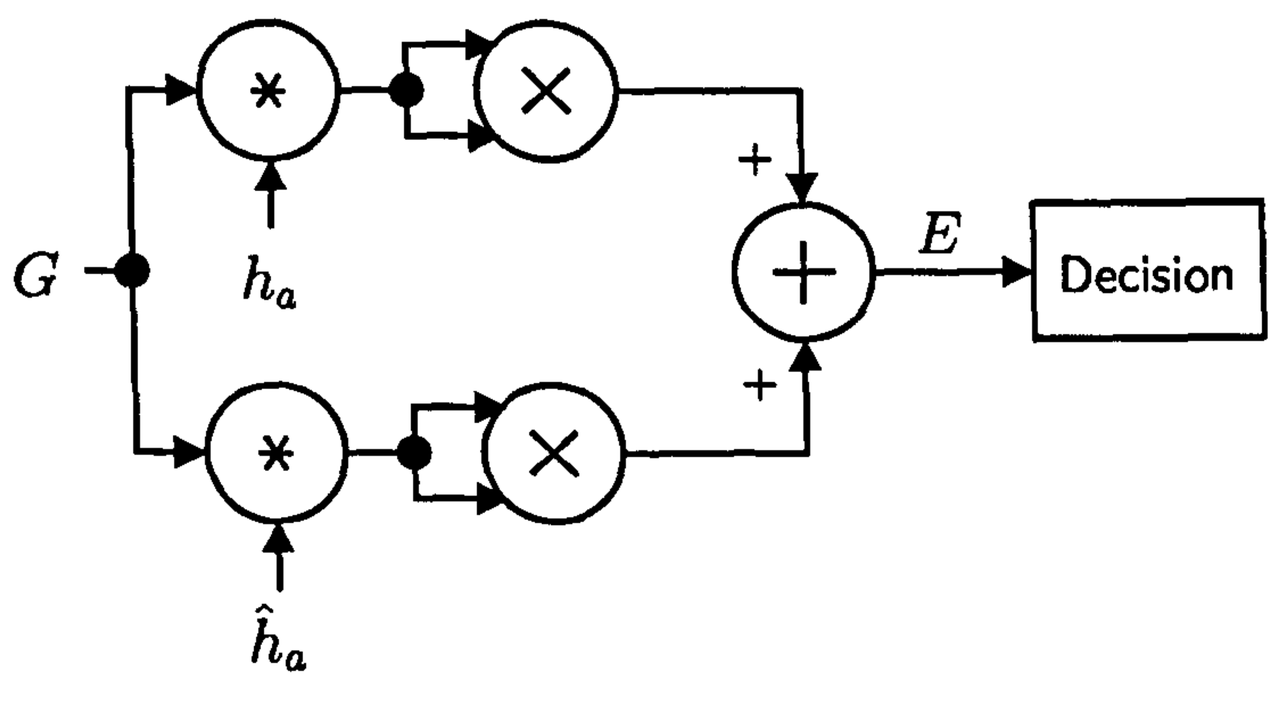


Figure 4.4: Flow diagram for the squared-envelope detector. (The lower branch depicts the quadrature-phase signal as the convolution of the real input process, G, with the Hilbert transform (Whalen, 1971, Chapter 3) of the analysis filter impulse response, denoted $\hat{h}_a$.)

4.2 Squared Envelope Detector

4.2.1 Overview

A squared-envelope detector samples the signal envelope, squares it, and performs a hypothesis test to decide whether a signal is present in the noise. This type of detector provides an ideal baseline against which to compare the performance of the interval receivers presented in later sections. The essential narrowband passive sonar comprises a squared-magnitude DFT with an adjustable threshold, which is simply a bank of squared-envelope detectors operating within narrowband channels. Accordingly, the conclusions we draw about narrowband envelope detection below apply equally well to DFT detection at an equivalent resolution.

The test statistic is computed by filtering the received signal through the analysis filter and measuring a single sample of the squared-envelope, as illustrated by the flow diagram in Figure 4.4. The envelope is governed by a random variable, E; we denote an observation using e. Assuming uniform prior probabilities, that is, that the probability of the target signal's presence is equal to that of its absence, the decision rule satisfying the minimum error criterion is

$$\text{choose } H_1 \text{ iff } \frac{p_E(e \mid H_1)}{p_E(e \mid H_0)} > 1, \text{ otherwise choose } H_0. \tag{4.7}$$

In summary, the test statistic, e, is first computed by squaring and adding the in-phase and quadrature components of the narrowband filter output, and then supplied to the likelihood test in (4.7). This is shown schematically in Figure 4.4. The next step is to obtain the conditional probability density functions, $p_E(e \mid H_j)$, which make up the likelihood ratio.

4.2.2 Probability Density Functions

It is well known that the probability density function governing the squared-envelope of a zero mean, wide-sense stationary Gaussian process is that of the exponential distribution (Whalen, 1971). This p.d.f. has the form

$$p_E(e;\sigma^2) = \frac{1}{2\sigma^2}\exp\left(\frac{e}{-2\sigma^2}\right), \tag{4.8}$$

in which the process variance, σ^2, is the only parameter. Let the variance of the random process X conditioned on H_j be denoted σ_j^2. The likelihood ratio simplifies to the expression

$$\lambda(e) = \frac{\sigma_0^2}{\sigma_1^2}\exp\left(\frac{e}{2}\left[\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right]\right). \tag{4.9}$$

Once the process variances have been determined, the likelihood ratio (4.9) is fully parameterised, and the detection process can begin. The random processes under consideration in this section are wide-sense stationary, so their variances are equivalent to the autocovariance functions evaluated at zero, i.e.,

$$\sigma_j^2 \equiv E\{X^2 \mid H_j\} \equiv \gamma_j[0].$$

The $\gamma_j[0]$ can be found by application of the rules for determining the autocovariance functions and squared-magnitude responses for systems of linear filters combined in series or in parallel (Whalen, 1971).

4.2.3 Setting up the Experiments

To predict the performance of the squared-envelope detector, we must determine the decision regions and then appropriately integrate the conditional probability density functions in those regions to find the probability of false alarm and false dismissal. The performance metric adopted in this chapter is the average of these latter quantities: the probability of error. The decision region $\mathcal{R}_0$ we define as the set of all envelope measurements which lead the detector to choose H_0, i.e., to decide that only noise has been received,

$$\mathcal{R}_0 = \{e \in \mathbb{R} : \lambda(e) \leq 1, e \geq 0\}; \tag{4.10}$$

similarly, we define $\mathcal{R}_1$ as the set containing all the values for the envelope for which the detector chooses H_1,

$$\mathcal{R}_1 = \{e \in \mathbb{R} : \lambda(e) > 1, e \geq 0\}. \tag{4.11}$$

From (4.10), we can specify $\mathcal{R}_0$ for the squared-envelope detector, as any real value e satisfying the following inequalities

$$\frac{\sigma_0^2}{\sigma_1^2}\exp\left(\frac{e}{2}\left[\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right]\right) \leq 1, \tag{4.12}$$

$$e \geq 0, \tag{4.13}$$

which, together, can be re-arranged into

$$0 \leq e \leq \frac{2\ln\sigma_1^2 - 2\ln\sigma_0^2}{1/\sigma_0^2 - 1/\sigma_1^2} \equiv \epsilon. \tag{4.14}$$

This indicates a single decision boundary at $e = \epsilon$. All the remaining possible values of the envelope, namely those which satisify $e > \epsilon$, constitute the decision region R_1. Consequently, the probabilities of false alarm and false dismissal evaluate to

$$P(D_1 \mid H_0) \triangleq \int_{\epsilon}^{\infty} p_E(e \mid H_0)de = \exp\left(\frac{\epsilon}{-2\sigma_0^2}\right) \tag{4.15}$$

$$P(D_0 \mid H_1) \triangleq \int_{0}^{\epsilon} p_E(e \mid H_1)de = 1 - \exp\left(\frac{\epsilon}{-2\sigma_1^2}\right), \tag{4.16}$$

respectively. Under the assumption of uniform priors, the probability of error is given by the average of (4.15) and (4.16), and is a function of σ_0^2 and σ_1^2 alone.

4.2.4 Experimental Results and Analysis

How does the detector's performance vary with SNR? The top three plots in Figure 4.5 show that the probability of error decreases monotonically as the signal-to-noise ratio increases, regardless of whether the signal is placed at the band centre (left-hand plot) or the band-edge (right-hand plot).

What is the effect of displacing the signal from the band centre? As the magnitude response of the analysis filter is bell-shaped, shifting the frequency of the signal further away from the centre of the analysis band attenuates the signal but passes the same noise power. The squared-envelope detector relies exclusively on power, so displacing the signal from the band centre lowers the post-analysis SNR and causes the probability of error to rise. This is revealed implicitly in the top three plots of Figure 4.5, where the probability of error is larger for greater displacements; and explicitly in the bottom right-hand plot of Figure 4.5, which plots the probability of error against signal frequency (f_c). Recall that in each case, the detector has been calibrated to account for the change in frequency; in other words, these results are optimal, given full knowledge of the signal and noise conditions.

Does the detection performance relate to the absolute frequency of the signal? No. The detection performance of the squared-envelope detector is only negligibly affected by a wholesale shift in frequency, provided that the analysis band and signal coincide and are shifted in frequency the same amount. This is confirmed in the bottom right-hand panel of Figure 4.5, which plots a constant probability of error with respect to band and signal frequency.

Is it possible to predict the performance of the detector? Yes. The performance of the quadrature detector is well-documented (Dawe, 1997; Burdic, 1984; Whalen, 1971) and straight-forwardly predicted from the exponential probability density functions described in Section 4.2.3. In all five graphs of Figure 4.5, the analytical predictions, plotted as solid lines, provide a close match to the empirical measurements, which are superimposed as markers.

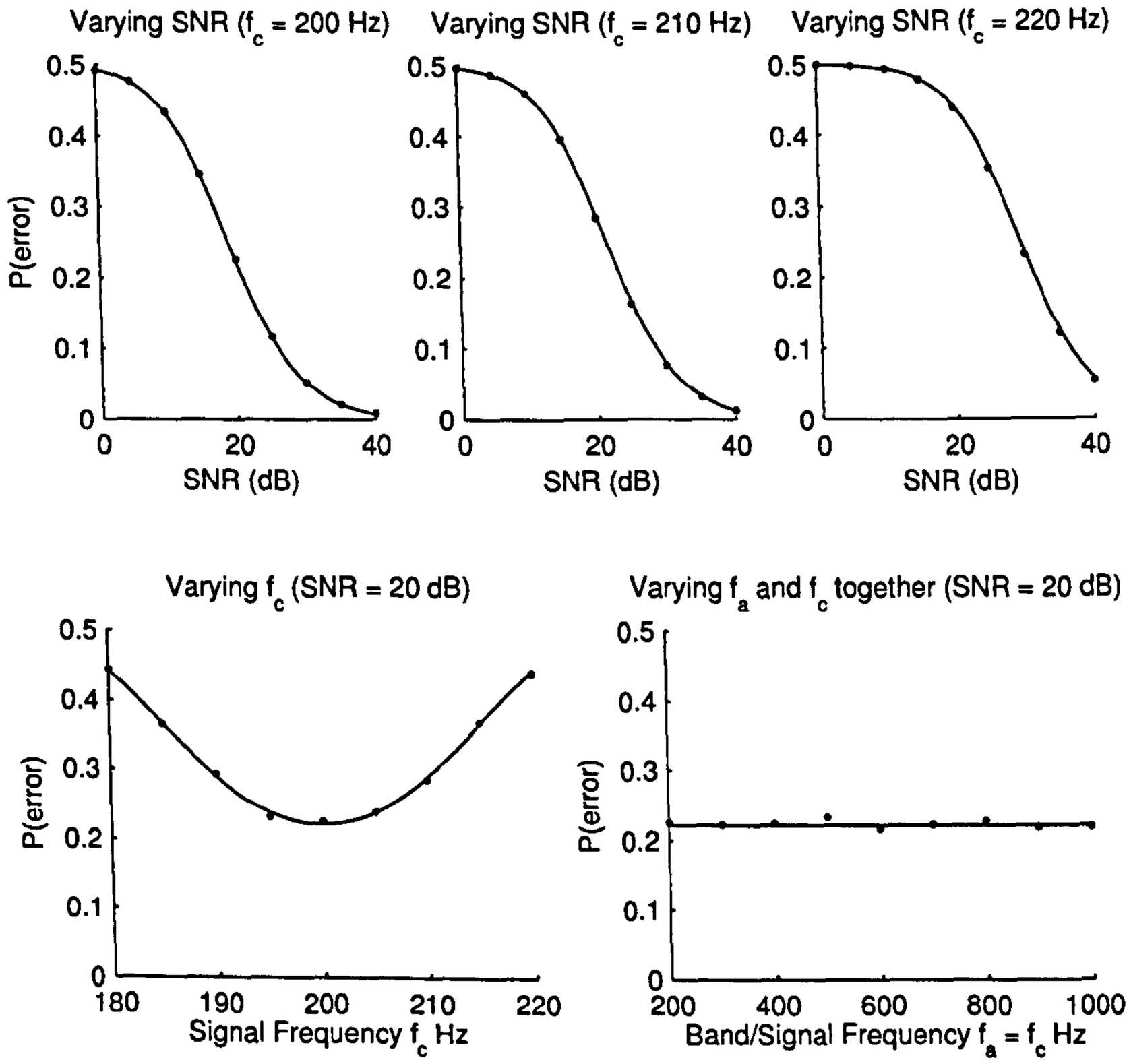


Figure 4.5: Probability of error in the squared-envelope detector: predicted values for squared-envelope detector (solid line); observed values for squared-envelope detector (solid circles •).

4.3 Sampled Interval Detector

4.3.1 Overview

Having established the squared-envelope detector as a baseline, we now turn to the first of the three interval detectors discussed in this chapter: the *sampled interval detector* (SID). The sampled interval detector extracts an interval test statistic, designated i, by differencing two consecutive zero crossing sample times in the output of a narrowband filter. This test statistic is then submitted to the minimum-error decision rule

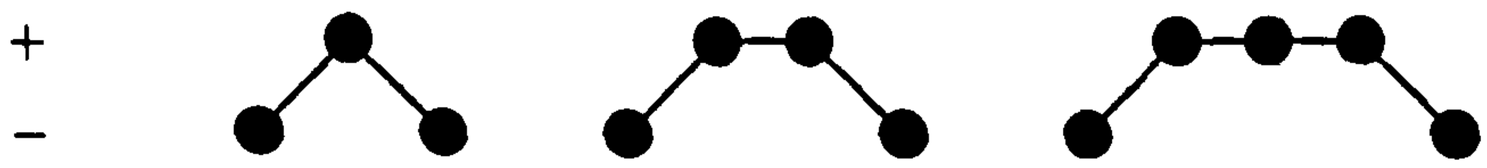
$$\text{choose } H_1 \text{ iff } \frac{p_I[i \mid H_1]}{p_I[i \mid H_0]} > 1, \text{ otherwise choose } H_0. \tag{4.17}$$

Construction of the sampled interval detector falls into two stages: formalising how a zero crossing interval is extracted and deriving the probability density functions that appear in the likelihood ratio. Once these tasks are complete, we shall be ready to compare the performance of the sampled interval detector against that of the squared envelope detector from the previous section.

The Interval Statistic

A zero crossing interval statistic is obtained whenever a zero crossing detector fires, so detection decisions coincide with the zero crossing times. (Contrast this with the squared-envelope detector, for which a test statistic is available at every sample.) The interval is computed by differencing the current sample time with that of the previous zero crossing, which is held in a buffer. Figure 4.6 provides a block diagram for the sampled interval detector. A useful block to define is the '`hold` block', shown in 4.6A, which serves as a simple memory in all three interval detectors. If the input on `IN1` is zero, then the `hold` unit retains the currently stored value. When `IN1` is non-zero, then the switch inside is 'up', and the memory is updated with the input on `IN0`. The sampled interval detector combines ZC and `hold` blocks to compute the zero crossing interval shown in 4.6B.

A block diagram is not the only interpretation of a zero crossing interval statistic. Further clarification is provided by a some examples of binary waveforms such as those shown below.



$i = 1$ $i = 2$ $i = 3$

The question of how the value of the test statistic i relates to the number of samples in the waveform has an intuitive answer. Because the zero crossings are sampled, in the absence of additional information, it is natural to place the zero crossing time halfway between the samples; differencing these times then gives rise to the interval statistics shown. Note that i is always an integer greater than or equal to one.

Figure 4.6: A. `Hold` block for buffering a value; B. block diagram for the sampled interval detector.

4.3.2 Probability Density Functions

The zero crossing interval i is derived from the samples of a random process and hence is governed itself by a random variable, labelled I. In this section, we determine the probability density for I given a particular hypothesis, $p_I[i \mid H_j]$, as a key step towards completing the decision rule in (4.17), or, equivalently, appreciating the inner working of the 'decision' block in Figure 4.6. The method employed to do this continues in the tradition of Kedem (1986), following Rice (1944), in determining the probability of a pattern of zero crossings by integrating the p.d.f.s governing the samples of an evolving process.

Interval Probability in a Wide-sense Stationary Process

Let us assume that the output of the analysis filter, $x[n]$, is a zero mean, wide-sense stationary random process and predict the probability of a pattern of sign changes given by

$$P(x_n < 0, x_{n-1} \geq 0, x_{n-k-1} < 0). \tag{4.18}$$

This pattern corresponds to observing a zero crossing from positive to negative at sample time n, as well as a negative sample $k+1$ samples earlier. Given that the process

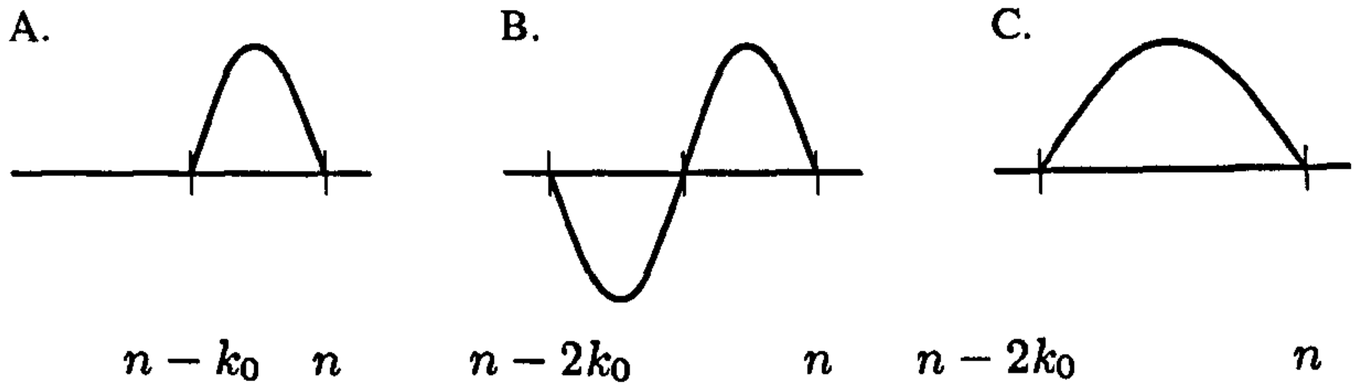
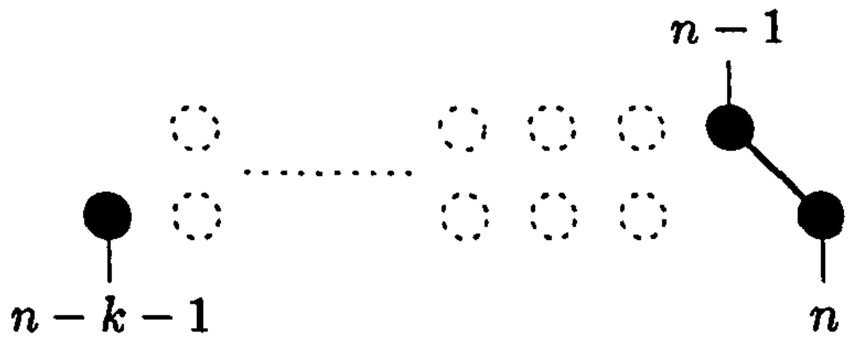


Figure 4.7: A) constrain the shortest possible interval to be k_0 samples; B) multiple crossings can only be obtained beyond $2k_0$; C) constrain the longest possible interval to be $2k_0$ samples.

is wide-sense stationary, we can assume that the probability of observing this pattern (4.18) is the same, regardless of n. This random event is illustrated in the sketch below: samples of known sign are shown as filled circles; unknown samples are open circles.



If such a pattern is observed, then evidently at least one zero crossing must have occurred between $n-1$ and $n-k-1$ to effect the change of sign. More specifically, we can infer that there are an odd number of crossings in the unknown samples. Because the number of sign changes is unknown, the evidence from just three samples is ambiguous. We are not in a position to assign the probability to one long interval, as these samples could also indicate many short intervals.

To remedy this, we impose a constraint upon the random variable,

$$k_0 < I, \tag{4.19}$$

effectively stating that intervals shorter than or equal to k_0 are impossible. With (4.19) in place, it is certain for all $k < 2k_0$ that *at most* one zero crossing is present. The reasoning behind this is as follows: taking the shortest interval possible and placing two or more in sequence always extends to a point equal to or beyond $2k_0$, as shown in Figure 4.7B.

Now we can write:

$$P(I_s \leq k) = \begin{cases} 2P(x_n < 0, x_{n-1} \geq 0, x_{n-k-1} < 0) & 0 < k < 2k_0 \\ 0 & k \leq 0 \\ \text{not determined} & k \geq 2k_0. \end{cases} \tag{4.20}$$

where $P(I_s \leq k)$ denotes the probability of an interval not exceeding length k on any given sample, and the probability mass is doubled to account for the same interval with the zero crossings in the opposite direction.

The probability that an interval longer than $2k_0$ samples is observed remains difficult to determine, because it could still contain multiple crossings. A direct solution to this problem is to add an additional constraint, placing an *upper* limit on interval duration, i.e.,

$$k_0 < I < 2k_0. \tag{4.21}$$

Let us take a moment to reflect on these constraints. First, observing an interval shorter than k_0 at any given sample is impossible. Second, because all intervals must be shorter than $2k_0$, it follows that the probability of observing an interval shorter than $2k_0$ on a given sample is the same as the probability of observing *any* interval, which, in turn, is as probable as a zero crossing. Putting this together,

$$P(I_s \leq k) = \begin{cases} 2P(x_n < 0, x_{n-1} \geq 0, x_{n-k-1} < 0) & k_0 < k < 2k_0 \\ 0 & k \leq k_0 \\ P(I_s < 2k_0) = P(C) & k \geq 2k_0. \end{cases} \tag{4.22}$$

(4.22) expresses the probability that: a zero crossing occurs on a given sample, and a second zero crossing occurs at most k samples earlier. The final step is to find the probability that an interval has a certain length, given that an interval has been received. Interval events coincide with zero crossings events, so

$$P(C)P(I \leq k) \equiv P(I_s \leq k). \tag{4.23}$$

Hence, the cumulative distribution function for I is shown to be

$$P(I \leq k) = \begin{cases} \dfrac{2P(x_n < 0, x_{n-1} \geq 0, x_{n-k-1} < 0)}{P(C)} & k_0 < k < 2k_0 \\ 0 & k \leq k_0 \\ 1 & k \geq 2k_0. \end{cases} \tag{4.24}$$

Interval Probability in a Gaussian Process

By placing suitable constraints on the duration of intervals, we obtained a cumulative distribution function for a general wide-sense stationary process. In order to determine the distribution of zero crossing intervals for a Gaussian process in particular, we must replace in (4.24) the quantities $P(C)$ and

$$2P(x_n < 0, x_{n-1} \geq 0, x_{n-k-1} < 0) \tag{4.25}$$

with expressions specific to joint Gaussian density functions. It has already been stated in (1.12) that, for a wide-sense stationary, zero mean Gaussian process with autocorrelation function $\rho[k]$,

$$P(C) = \frac{1}{2} - \frac{1}{\pi} \sin^{-1} \rho[1]. \tag{4.26}$$

All that remains, then, is to evaluate (4.25). A useful procedure adopted by Kedem (1986) defines an indicator function for a sample

$$d_n = \begin{cases} 1 & x_n \geq 0 \\ 0 & x_n < 0 \end{cases} \tag{4.27}$$

so that the probability $P(x_n \geq 0)$ can be written as an expectation of the indicator function, i.e., $E\{d_n\}$. Using this notation, we can specify (4.25) as

$$\begin{aligned} & 2P(x_n < 0, x_{n-1} \geq 0, x_{n-k-1} < 0) \\ &= P(x_n < 0, x_{n-1} \geq 0, x_{n-k-1} < 0) \\ &\quad + P(x_n \geq 0, x_{n-1} < 0, x_{n-k-1} \geq 0) \qquad (4.28) \\ &= E\{(1-d_n)d_{n-1}(1-d_{n-k-1})\} + E\{d_n(1-d_{n-1})d_{n-k-1}\} \qquad (4.29) \\ &= E\{d_n\} - E\{d_n d_{n-1}\} - E\{d_n d_{n-k}\} + E\{d_n d_{n-k-1}\}. \qquad (4.30) \end{aligned}$$

Notice that all the terms in (4.30) express two-dimensional orthant probabilities, with the exception of the first, which evaluates to $\frac{1}{2}$. Replacing each of the orthant probabilities with an appropriate expression of the form

$$\frac{1}{4} + \frac{1}{2\pi}\sin^{-1}\rho[\cdot]$$

(Kedem, 1980), gives

$$\begin{aligned} & 2P(x_n < 0, x_{n-1} \geq 0, x_{n-k-1} < 0) \\ &= \frac{1}{2} - \left[\frac{1}{4} + \frac{1}{2\pi}\sin^{-1}\rho[1]\right] \\ &\quad - \left[\frac{1}{4} + \frac{1}{2\pi}\sin^{-1}\rho[k]\right] + \left[\frac{1}{4} + \frac{1}{2\pi}\sin^{-1}\rho[k+1]\right] \qquad (4.31) \\ &= \frac{1}{4} + \frac{1}{2\pi}\left(\sin^{-1}\rho[k+1] - \sin^{-1}\rho[1] - \sin^{-1}\rho[k]\right). \qquad (4.32) \end{aligned}$$

Finally, placing (4.26) and (4.30) into (4.24) and cancelling terms results in an expression for the cumulative distribution function of the intervals for a Gaussian random process, exclusively in terms of its autocorrelation function:

$$P(I \leq k) = \begin{cases} \dfrac{1}{2} + \dfrac{\sin^{-1}\rho[k+1] - \sin^{-1}\rho[k]}{\pi - 2\sin^{-1}\rho[1]} & k_0 < k < 2k_0 \\ 0 & k \leq k_0 \\ 1 & k \geq 2k_0. \end{cases} \qquad (4.33)$$

The probability density function for I is obtained by differencing (4.33)

$$p_I[i] = P(I \leq i) - P(I < i), \qquad (4.34)$$

to give

$$p_I[i] = \begin{cases} \dfrac{\sin^{-1}\rho[i+1] - 2\sin^{-1}\rho[i] + \sin^{-1}\rho[i-1]}{\pi - 2\sin^{-1}\rho[1]} & k_0 < i < 2k_0 \\ 0 & \text{otherwise.} \end{cases} \qquad (4.35)$$

We are now in a position to compute the autocorrelation function for a wide-sense stationary (Gaussian) process, using the techniques described at the beginning of this chapter, and, by applying (4.35), to find the probability density function governing the intervals.

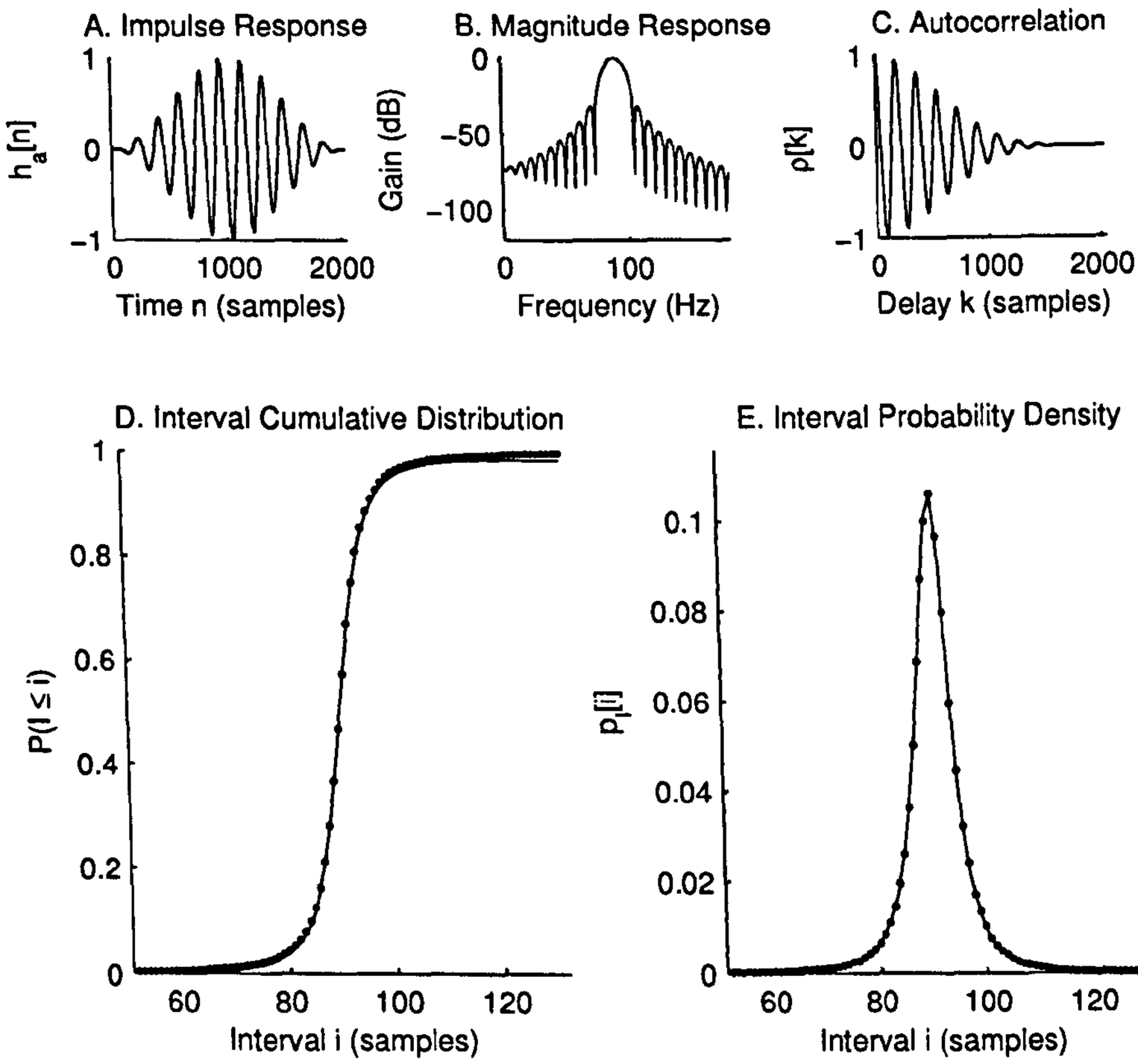


Figure 4.8: Stages in computing the interval probability density function. A) System impulse response; B) squared-magnitude response (dB attenuation with respect to the peak); C) autocorrelation function; D) analytical (solid line) and empirical (solid circle) c.d.f.; E) analytical (solid line) and empirical (solid circle) p.d.f..

Comparing the Analytical and Empirical Distributions

The analytical probability density function for I is first determined by combining the steps described earlier and then compared to a histogram formed by random trials. The comparisons that follow are not intended to validate the preceding working formally, but rather to provide a collection of examples as a visual aid.

The procedure involves convolving white Gaussian noise with a linear filter and comparing the analytical and empirical distribution and density functions. Our earlier working demands that the intervals be confined to the range $k_0 < I < 2k_0$ for some k_0, according to (4.21). For now we shall rely on the intuition that narrow band-pass filtered noise centered at f_a Hz generates intervals in a correspondingly narrow range surrounding

$$\text{mean interval} \approx \frac{1}{2f_a}, \text{ seconds.} \tag{4.36}$$

A straight-forward method of constructing the impulse response for a finite impulse response (FIR) filter is to apply a tapered window, $w[n]$, to a sinusoid at the filter centre frequency. The window length and bandwidth are inversely related, and the choice of window dictates the filter shape in the frequency domain. This gives rise to an impulse response whose general form is

$$h_a[n] = w[n]\sin(2\pi f_a n/f_s). \tag{4.37}$$

The following example employs $f_a = 90$ Hz as the centre frequency, $f_s = 16384$ Hz as the sample rate, and the Hann window (Oppenheim and Schafer, 1989) as $w_H[n]$, defined by

$$w_H[n] = \begin{cases} 0.5\left[1 - \cos\left(\dfrac{2\pi n}{N-1}\right)\right], & 0 \le n \le N-1 \\ 0 & \text{otherwise.} \end{cases} \tag{4.38}$$

where $N = 2048$ samples. The impulse response (4.38) is plotted in Figure 4.8A. The squared magnitude response of the filter is computed from its impulse response (using $2N$ samples to avoid aliasing), i.e.,

$$|\mathscr{H}_a[s]|^2 = \left|\sum_{n=0}^{2N-1} h_a[n]e^{-is\pi n/N}\right|^2, \; 0 \le s \le 2N-1 \tag{4.39}$$

and is plotted in Figure 4.8B on a logarithmic scale. If the white noise input to the filter has power σ^2, then it follows that the power spectral density of the output process, here designated X, is

$$\mathscr{S}_X[s] = \sigma^2|\mathscr{H}_a[s]|^2 \tag{4.40}$$

from which, via the Wiener-Khinchin relations (Shanmugan and Breipohl, 1988), are obtained the autocovariance and autocorrelation functions:

$$\gamma_X[k] = \begin{cases} \dfrac{1}{2N}\displaystyle\sum_{s=0}^{2N-1} \mathscr{S}_X[s]e^{is\pi k/N} & |k| < N \\ 0 & \text{otherwise} \end{cases} \tag{4.41}$$

and

$$\rho_X[k] = \frac{\gamma_X[k]}{\gamma_X[0]} \tag{4.42}$$

respectively. The autocorrelation function $\rho_X[k]$ is shown in Figure 4.8C for positive values of k. As a final step, the interval cumulative distribution function and interval probability density function are found by placing $\rho_X[k]$ into (4.33) and (4.35), the results of which are plotted in Figures 4.8D and 4.8E as solid lines.

The empirical distribution function is simply generated by passing white Gaussian noise through the linear filter with the impulse response $h_a[n]$, recording the zero

crossing intervals in the output, and using the estimate

$$\hat{P}(I \leq k) = \frac{\text{number of intervals less than or equal to } k}{N} \tag{4.43}$$

where $\hat{P}(\cdot)$ denotes an estimated probability and N is the number intervals measured (Shanmugan and Breipohl, 1988). Intervals whose separation in time is shorter than the length of the impulse response are statistically dependent; however, the estimated distribution converges upon the true population distribution when N is sufficiently large.

Similarly, the probability density function may be estimated from the data itself

$$\hat{p}_I[i] = \frac{\text{number of intervals equal to } i}{N}, \tag{4.44}$$

or found by differencing the empirical cumulative distribution. Figures 4.8D and 4.8E plot the empirical cumulative distribution and probability density function obtained from 200,000 measurements using solid circles.

The analytical and empirical distributions align closely at all points except the tail of the distribution corresponding to long intervals. The departure is most clearly evident in the cumulative distribution function for intervals around 120–130 samples, in which the analytical function appears to form a plateau prematurely at approximately 0.96, whilst the empirical version approaches one. Aside from this discrepancy, the cause of which will be examined in the next section, the predicted and observed density functions appear to be a close match. To provide a broader perspective, the analytical and empirical interval p.d.f.s were obtained by the same procedure for a variety of linear systems with a narrow band-pass frequency response. The results included in Figure 4.9 indicate that the proposed method for obtaining the analytical density function applies to other filter shapes, centre frequencies and bandwidths.

4.3.3 Interval Aliasing

In the previous section, we arrived at the cumulative distribution function for the intervals for a Gaussian process by two distinct approaches. The first was an analytical solution obtained from the autocorrelation function of the process; the second was an empirical solution found by synthesising the Gaussian process and measuring its intervals. In some cases, the distribution functions were noticed to differ slightly in the tail, an artifact which is most evident in Figure 4.8D. In the following section, we investigate cause of this discrepancy.

Recall that, when deriving the probability of an interval, we considered the probability that a zero crossing is preceded by a sign change $k-1$ samples earlier. This sign change is sufficient to indicate that an interval shorter than, or equal to k samples has occurred; indeed, at first, it might appear that we have found the cumulative distribution function for I. Such an analysis would be mistaken, however. The sign change could signal the presence of many intervals, not one—and we have no information about the intermediate samples by which to judge.

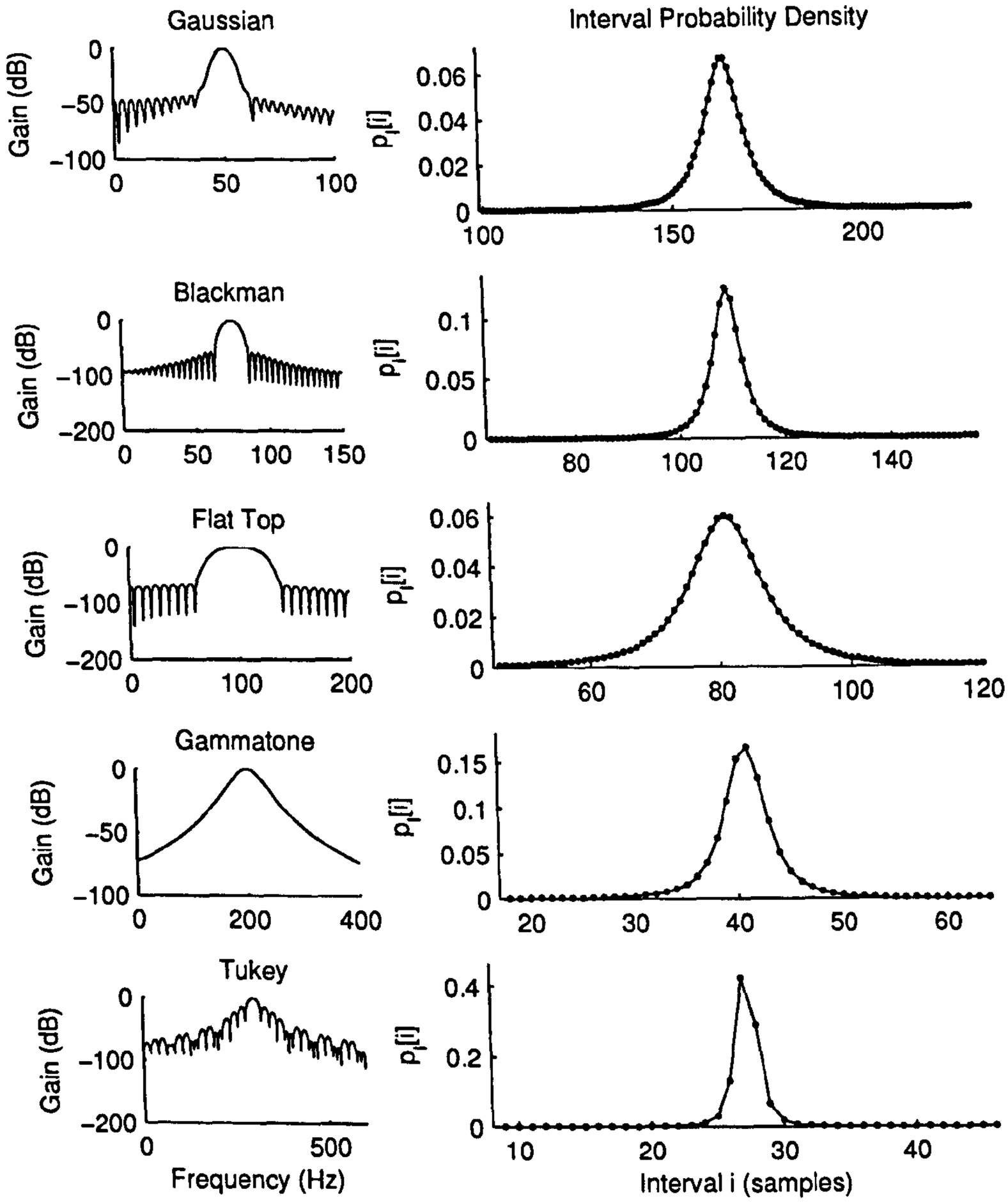


Figure 4.9: Interval probability density functions for various band-pass filter shapes. The log-magnitude responses for each filter, normalised so that peak response has unit gain, are plotted on the left. The interval p.d.f. associated with each filter is shown on the right: the solid line shows the analytical p.d.f., and the solid markers plot the histogram. All the window functions are provided by the MATLAB signal processing toolbox, with the exception of the gammatone filter, whose impulse response is $h_a[n] = n^4 e^{-n/70} \sin(2\pi f_a n / f_s)$.

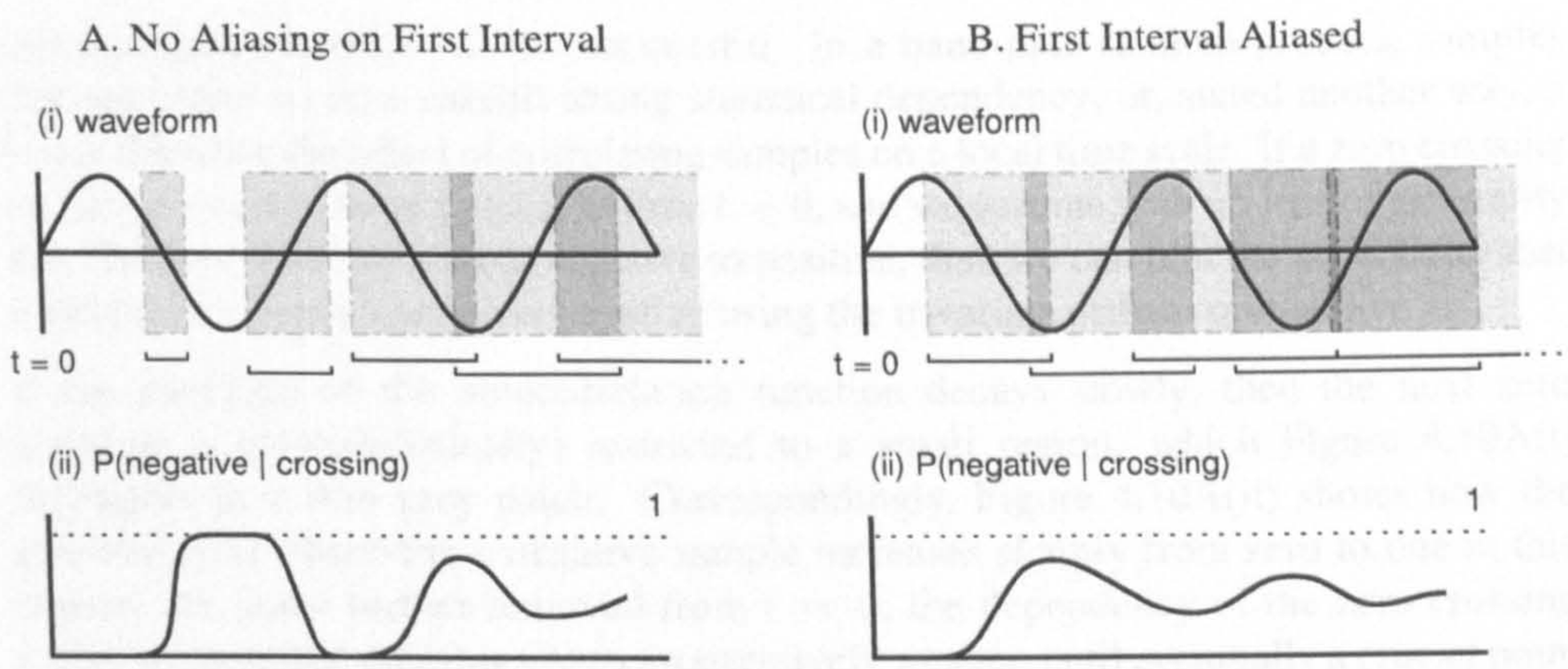
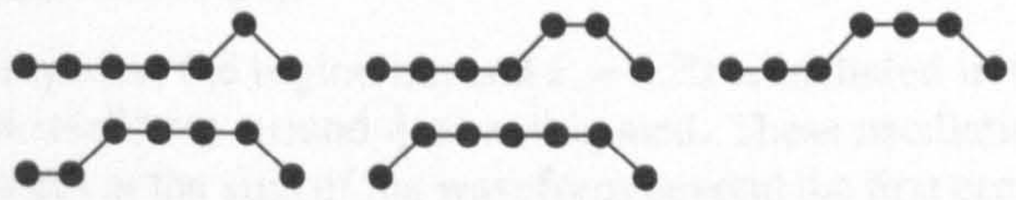
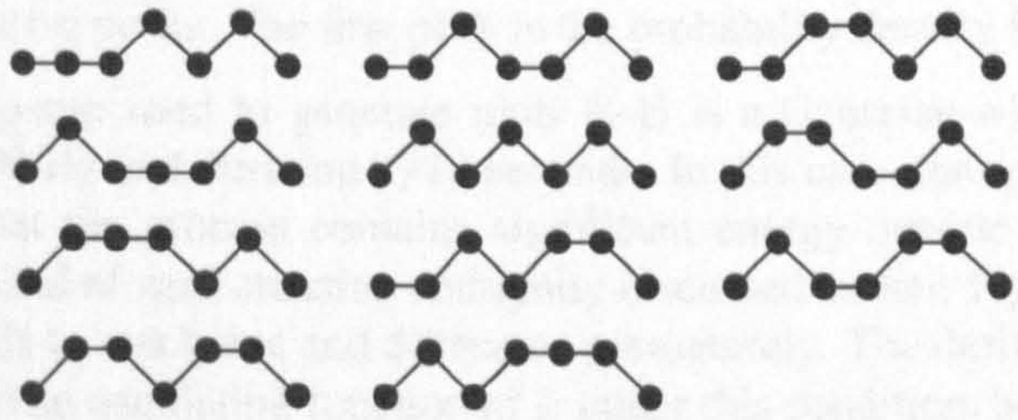


Figure 4.10: Ambiguity in zero crossings. A) the third interval is aliased; B) the first interval is aliased. *Top plots:* white regions are guaranteed to be zero crossing-free, grey regions indicate where a zero crossing might fall, and overlapping regions are highlighted in darker shades. *Bottom plots:* the probability of a negative sample waxes and wanes according to the zero crossing regions.

Let us take the example $k = 5$; that is, at a particular sample, a zero crossing has been observed and the reversal of sign occurs six samples earlier. There are five ways to interpret the samples which yield a single interval (i.e., $i \in \{1, 2, 3, 4, 5\}$). These are sketched below.



There are also eleven alternative interpretations which include multiple intervals.



To prevent the multiple interval ambiguity from arising, we introduce constraints, limiting the intervals to the range $k_0 < I < 2k_0$, for some k_0. In regard to the example above, by enforcing $3 < I < 6$, no sample functions in the second collection can ever emerge. The signal will contain only the intervals $i \in \{4, 5\}$ and their probability can be determined unambiguously using the 'three-sample' approach.

It is useful to consider the waveforms (or, more properly, *sample functions*) that are typical of a band-pass random process, in order to gain an insight into whether the

three-sample approach will be successful. In a band-pass random process, samples that are closer in time exhibit strong statistical dependency; or, stated another way, a linear filter has the effect of correlating samples on a local time scale. If a zero crossing occurs in a narrowband process at time $t = 0$, and we assume with no loss of generality that the zero crossing is from negative to positive, then we can plot the probability that a particular sample later on is negative using the trivariate orthant probability.

If the envelope of the autocorrelation function decays slowly, then the next zero crossing is (probabilistically) restricted to a small region, which Figure 4.10A(i) highlights as a thin grey patch. Correspondingly, Figure 4.10A(ii) shows how the probability of observing a negative sample increases sharply from zero to one in this region. At times further removed from $t = 0$, the dependency of the zero crossing times on the initial crossing grows progressively weaker, until eventually a crucial point is reached at which the function $P(x_\tau < 0)$ no longer consists of sharp transitions between zero and one, but takes the form a decaying, oscillatory function. Figures B(i) and B(ii) illustrate a more severe form of this effect, in which the first zero crossing may be mistaken for the second.

With these principles in mind, one can observe the detrimental effect of zero crossing ambiguity upon the analytical and empirical interval c.d.f. and p.d.f. of a synthesised random process, as shown in Figure 4.11. The impulse response used to generate plots A–D is a Gaussian-windowed sinusoid with frequency 90 Hz and duration $1/8$ seconds. The spectral bandwidth is sufficiently narrow that almost all ambiguity in zero crossings is suppressed. The analytical cumulative distribution function (C) is well-formed within the distribution's support, 60–120 samples, and matches the empirically-derived version (A).

For illustrative purposes, the region beyond $i = 120$ is included in the plot, at which point the function oscillates around $\frac{1}{2}$ as anticipated. These oscillations correspond to probabilistic changes in the sign of the waveform beyond the first crossing. Informally, the function relates 'probably positive', 'probably negative', and so forth. When i exceeds the length of the impulse response, the probabilities of a positive and negative sample are both $\frac{1}{2}$. In the function's derivative (D), the oscillations appear as alternating positive and negative peaks. The first peak is the probability density function.

The impulse response used to generate plots E–H is a Gaussian-windowed sinusoid with frequency 90 Hz and duration $1/16$ seconds. In this case, the spectral bandwidth is doubled so that the process contains significant energy outside the octave band, resulting in the kind of zero crossing ambiguity discussed earlier. Figure 4.11G shows that the c.d.f. fails to reach one and decreases prematurely. The derivative of the c.d.f. (H) is once again an oscillating function of i; under this condition, however, the peaks are somewhat broader. Consequently, the first negative peak merges with the first positive peak, so that the tail of the probability density function is negative, reflecting the change of slope in the cumulative distribution function.

It seems appropriate to name this undesirable effect *interval aliasing*, as it recalls the overlap seen in frequency domain aliasing when components with frequencies higher than the Nyquist frequency are manifested at lower frequencies. In the present context, aliasing refers to the probability associated with the next sign change being accounted

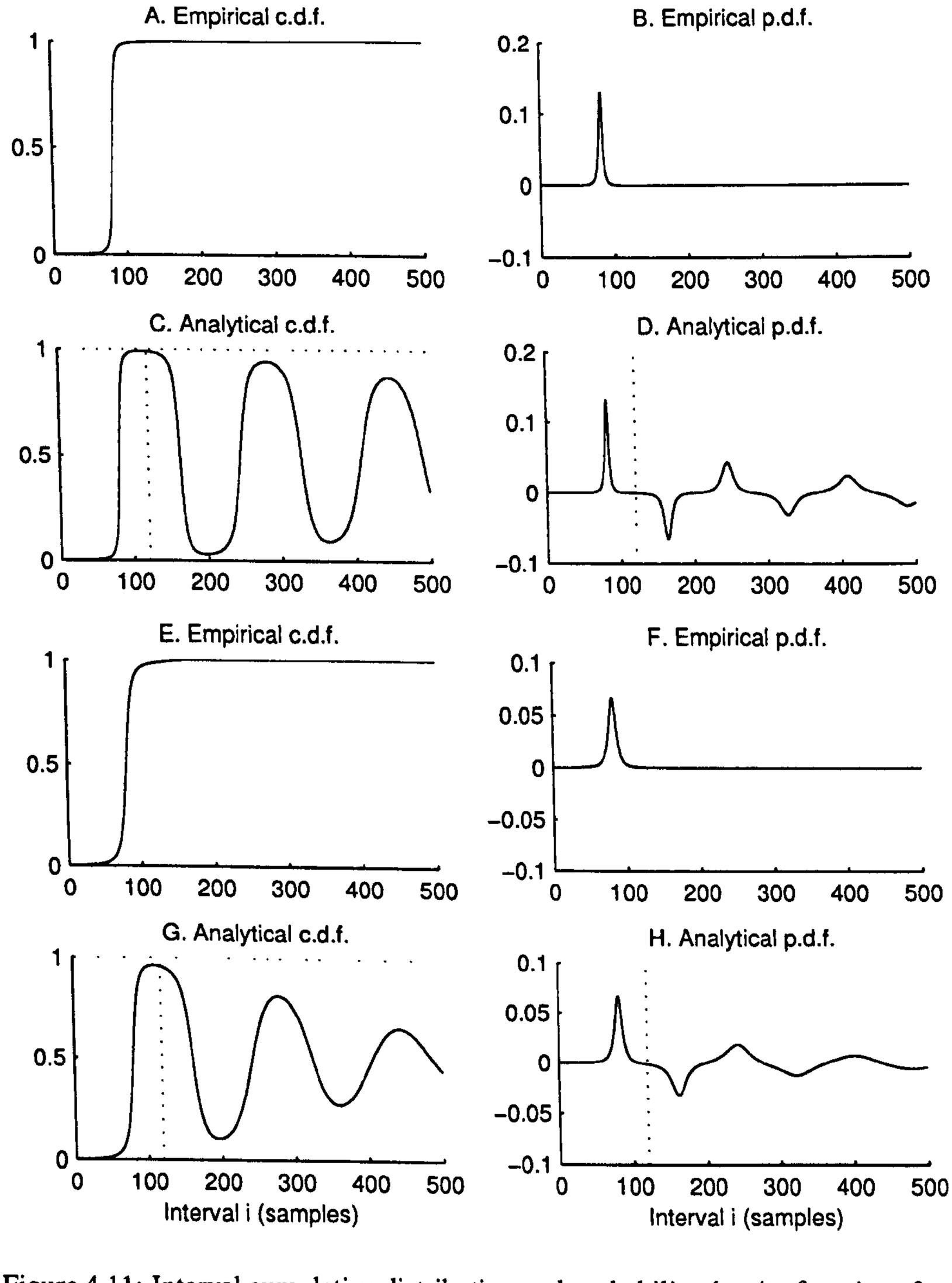


Figure 4.11: Interval cumulative distribution and probability density functions for two random processes, the first without (or with very little) aliasing (A–D), the second with an appreciable amount of aliasing (E–H). The dotted vertical lines show the upper limit of the interval p.d.f. support ($2k_0$). Notice that the c.d.f. in (G) does not reach one and the corresponding p.d.f. (H) contains negative values.

for along with the probability of later sign changes. As the preceding examples reveal, interval aliasing tends to lead to invalid analytical distributions, in particular, those with negative probability density. For the remainder of this thesis, we shall assume for convenience that an octave-band linear filter accomplishes the task of conditioning the zero crossings of a random process so that its zero crossing intervals satisfy a similar octave requirement.

4.3.4 Setting up the Experiments

Assuming that H_0 and H_1 are assigned equal prior probability, the sampled interval detector operates according to the rule

$$\text{choose } H_1 \text{ iff } \frac{p_I[i \mid H_1]}{p_I[i \mid H_0]} > 1, \text{ otherwise choose } H_0.$$

The likelihood functions in this case are the conditional probability density functions for i, which were determined in Section 4.3.2 for a Gaussian process, subject to reasonable bandwidth requirements. Such a detector is guaranteed to produce the fewest errors out of all the detectors that exclusively use i.

The probability of error for the sampled interval detector may be predicted for any particular task by summing the relevant portions of the density functions. Using D_0 and D_1 to respectively denote the event that H_0 and H_1 is chosen, the probability of a false alarm is

$$P(D_1 \mid H_0) = \sum_{i \in \mathcal{R}_1} p_I[i \mid H_0], \text{ where } \mathcal{R}_1 = \{i : p_I[i \mid H_0] < p_I[i \mid H_1]\} \quad (4.45)$$

and, similarly, the probability of a false dismissal is given by

$$P(D_0 \mid H_1) = \sum_{i \in \mathcal{R}_0} p_I[i \mid H_1], \text{ where } \mathcal{R}_0 = \{i : p_I[i \mid H_0] \geq p_I[i \mid H_1]\}. \quad (4.46)$$

The probability of error, assuming uniform priors, is then found by averaging the two types of error

$$P(\text{error}) = \frac{P(D_1 \mid H_0) + P(D_0 \mid H_1)}{2}. \quad (4.47)$$

4.3.5 Experimental Results and Analysis

The probability of error for the sampled interval detector and squared-envelope detector is shown for a variety of conditions in Figure 4.12. For each condition, the control parameters were set to those listed in Table 4.1 and a chosen parameter, or set of parameters, was varied according to the experiment design set out in Section 4.1.4.

How does the detector's performance vary with SNR? In the first experiment, the signal was placed at the centre of a 200 Hz analysis band and the effect upon the probability of error of changing the global signal-to-noise ratio was recorded for both the power and sampled interval detector. The results indicate that the interval detector

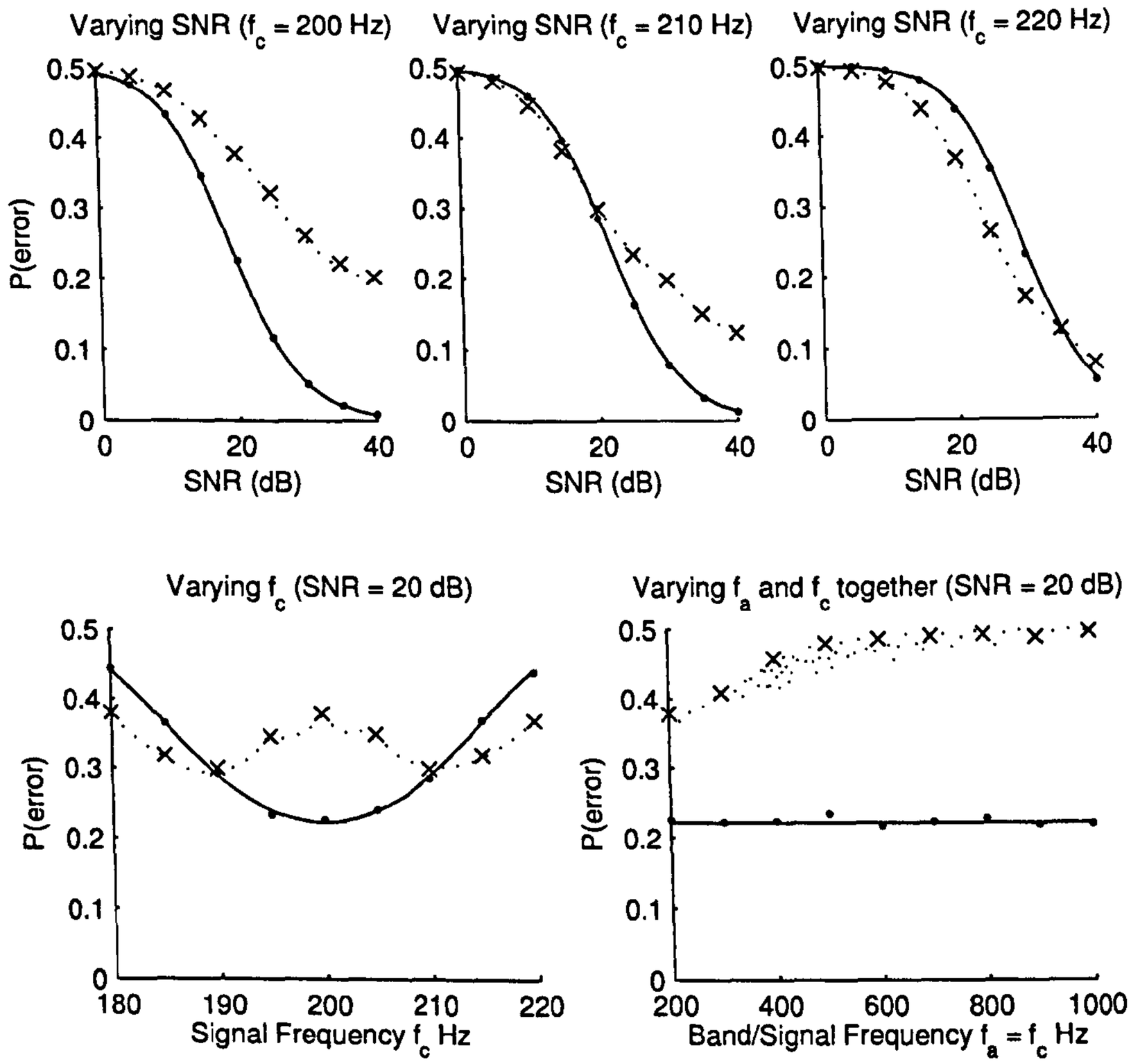


Figure 4.12: Probability of error in the sampled interval detector: predicted values for squared-envelope detector (solid line) and sampled interval detector (dotted line); observed values for squared-envelope detector (solid circles •) and sampled interval detector (crosses ×).

performs consistently worse than the squared-envelope detector, in that the probability of error is lower for the former than the latter in every condition. This difference is particularly pronounced when the SNR is high. For example, with the SNR set at 40 dB, the squared-envelope detector is almost free of errors, whereas the sampled interval detector has an error probability of 0.2, relating the misclassification of one in five intervals.

What is the effect of displacing the signal from the band centre? Subsequent experiments investigated the effect of moving the signal away from the band centre. The procedure described above was repeated with the signal placed at 210 Hz. As the top-centre plot in Figure 4.12 reveals, when the signal is displaced from the band centre by 10 Hz, the squared-envelope detector's performance worsens, that is, the probability of error is seen to increase. This result is to be anticipated: the analysis filter has a bell-

shaped magnitude response, so signals with frequencies removed from the band centre achieve a lower post-analysis SNR. The sampled interval detector, on the other hand, displays the opposite behaviour: displacing the signal *increases* the performance (i.e., lowers the probability of error). In fact, for signal-to-noise ratios lower than 20 dB, the interval detector starts to outperform the squared-envelope detector to a small extent. (The probability of error differs at most by about 0.01.)

To establish whether this trend persisted, another experiment was performed with the signal at 220 Hz—a displacement of 20 Hz. The top-right panel of Figure 4.12 plots the resulting error curves. Again, the probability of error for the squared-envelope detector increases at all SNRs. This time, however, the effect upon the interval detector is more complicated. In relation to the 210 Hz signal, the probability of error drops at high signal-to-noise ratios, e.g., 40 dB, whilst at low signal-to-noise ratios, e.g., 20 dB, the probability increases. Clearly then, there is no simple rule stating that increasing the displacement of the signal improves interval detection at all SNRs. (Such a result would be most counter-intuitive, suggesting that detection is optimal when the signal is severely attenuated.) Although the sampled interval detector shows a mixture of improvement and deterioration when compared to its own performance at 210 Hz, the improvement in performance of the interval detector over the squared-envelope detector is significant, with differences in error probability as large as 0.09.

These considerations give rise to the general question: How does the displacement of a signal from the band centre impact the probability of error? (For now, we shall assume that other parameters such as the band centre frequency and bandwidths remain fixed.) For the squared-envelope detector, with a unimodal analysis filter magnitude response, the relationship is evident: displacement increases the probability of error. As far as the sampled interval detector is concerned, the three experiments—with signal frequencies 200, 210 and 220 Hz—relate the signal displacement to probability of error at only three points: not enough to provide the complete picture. So, in the fourth experiment, the signal-to-noise ratio was held constant at 20 dB, and the probability of error was measured (and predicted) as a function of signal frequency. The curve that results is shown in the bottom left-hand plot of Figure 4.12.

The curve relating error to signal displacement appears to undergo two phases: the first corresponds to small displacements and a decrease in error; the second corresponds to larger displacements and an increase in error. The shape of the curve can be identified as arising from an interaction between two competing effects. To illustrate this requires returning to the conditional density functions involved. If the signal is placed at the centre of the band, then the conditional means for H_0 and H_1 coincide, in which case the higher-order moments, e.g., the variance, of the distributions must be called upon to distinguish the hypotheses. As the signal is moved away from the band centre, the means diverge and, as a result, the conditional p.d.f.s are better separated, leading to a performance improvement. At the same time, if the signal frequency becomes attenuated by the analysis filter, then the post-analysis SNR is lower, and the signal exerts a weaker effect over the p.d.f.. The degree of performance improvement at a particular frequency displacement depends on which of these two factors—the separation of conditional means or signal attenuation—dominates, as exemplified by the changing shape of the probability density functions in Figure 4.13.

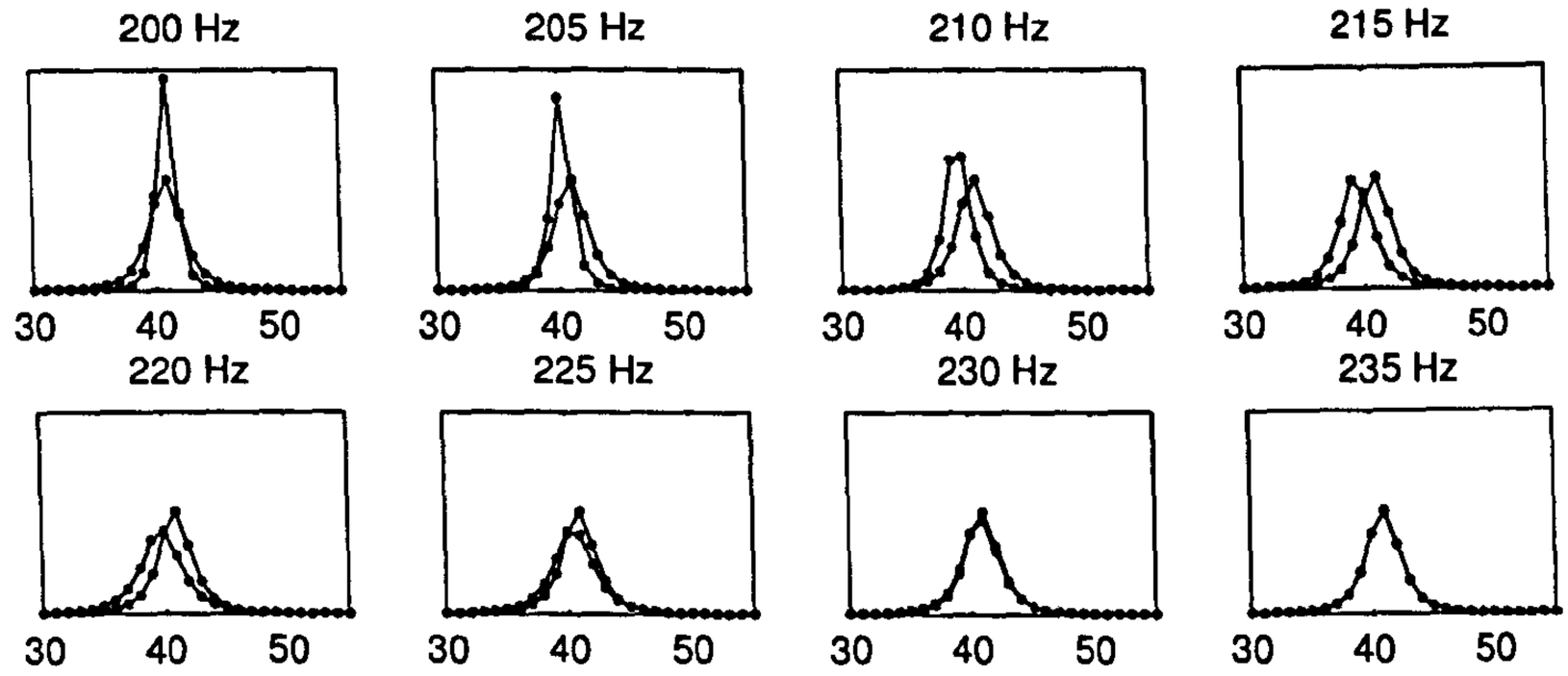


Figure 4.13: Probability density functions as signal frequency changes. Three distinct phases are identifiable: i) signal is centred on the band (200 Hz); ii) signal frequency increased, shorter intervals are received (205–215 Hz); iii) signal is attenuated and the p.d.f. for H_1 approaches that of H_0 (220 Hz and beyond).

Does the detection performance relate to the absolute frequency of the signal? The experiments reported so far describe the detection of a signal in noise within a band centred on 200 Hz, for various signal-to-noise ratios and signal frequencies. To determine whether the interval detector was effective at higher frequencies, further experiments were performed in which the signal and band frequency were varied together over the range 200 Hz–1 kHz. The bottom-right plot in Figure 4.12 confirms that the squared-envelope detector performs consistently: the expected outcome, given that the detections are made on the basis of power—unaffected by a shift in frequency. Meanwhile, the performance of the sampled interval detector is seen to degrade rapidly with increasing frequency. As before, this result can be interpreted by exploring the effect of the change in band frequency upon the conditional interval probability density functions. (See Figure 4.14.)

It is helpful to adopt the zero crossing intervals of the sinusoids corresponding to the analysis filter band edges (nominally, the -3 dB points) as a rough guide to the support of the interval p.d.f.. For instance, if the high-frequency cut-off falls at 500 Hz, then the shortest interval with non-zero probability would be estimated at 0.001 seconds. If we designate the filter bandwidth B Hz, the shortest and longest zero crossing intervals are

$$\frac{1}{2f_a \pm B}, \text{ seconds,}$$

respectively. The interval distribution range is then approximated by the difference in these two values:

$$\frac{1}{2f_a - B} - \frac{1}{2f_a + B} = \frac{2B}{4f_a^2 + B^2}. \tag{4.48}$$

Ostensibly, this analysis shows that the range and variance of the interval distribution is inversely proportional to a quadratic function of the band centre frequency, which implies that even modestly increasing the band frequency concentrates the probability mass in the discrete density functions over only a very few samples. Consequently, the

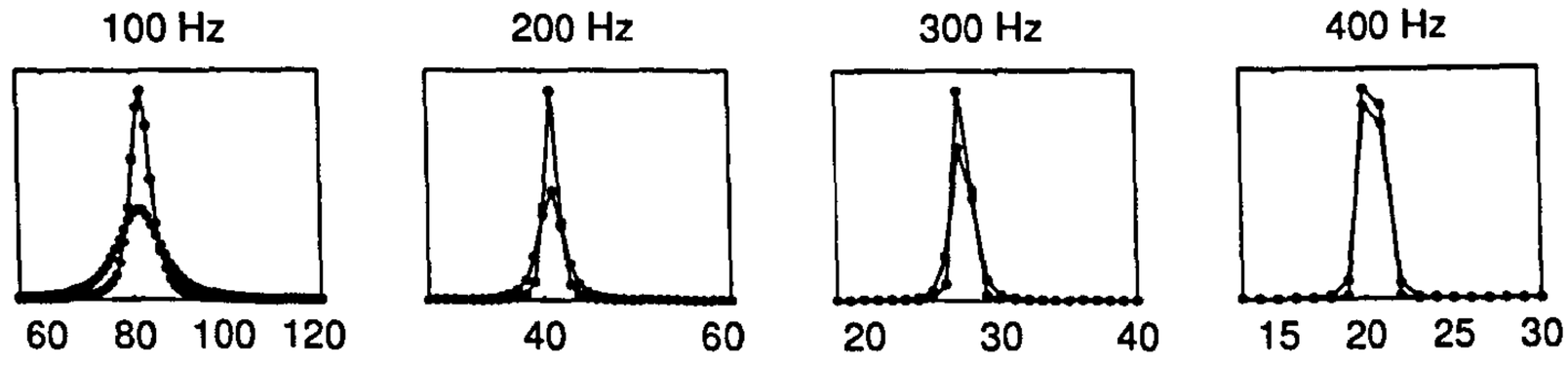


Figure 4.14: Varying the band and signal frequency affects the conditional probability density functions so that they are almost indistinguishable at higher frequencies.

performance of the sampled interval detector drastically diminishes as the band (and signal) are shifted up in frequency.

The resolution of the interval p.d.f. is based on the autocorrelation, so a direct approach to remedy this problem would be up-sampling the signal. In theory, this is a possibility; however, in order to maintain performance, doubling the band centre frequency would demand *quadrupling* the sample rate, and so on—an impractical suggestion. Another modification to counteract the effect of the increasing filter frequency could be to concomitantly increase the filter bandwidth with the centre frequency. Enforcing the relationship

$$QB = f_a$$

where Q is a positive constant, the approximation for the interval distribution range becomes

$$\frac{2B}{4f_a^2 + B^2} = \frac{2Q^2}{(4Q^2 + 1)f_a}. \tag{4.49}$$

Hence, for a constant-Q filter, the centre frequency and interval range are inversely proportional. Admittedly, this is less severe than the inverse-quadratic relationship arising from a constant bandwidth; nevertheless, using a constant-Q filter represents an unacceptable concession in terms of SNR, especially if a high-frequency, narrowband application is intended.

Is it possible to predict the performance of the detector? Inspecting Figure 4.12, the empirical probabilities of error match the predicted probabilities for both the squared-envelope and sampled interval detectors.

How do the results of the different detectors compare? Two main results have emerged from this study. First, the sampled interval detector commits fewer errors than the squared-envelope detector, when the signal is away from the centre of the analysis band; but the reverse is true when the signal is near the centre. Second, the sampled interval detector performs very poorly at modestly high frequencies, whereas the squared-envelope detector is naturally unaffected by a shift in frequency. It is suggested that the rounding of zero crossing times reduces the information available to the interval detector so that it fails to discriminate between signal and noise.

4.4 Continuous Interval Detector

4.4.1 Overview

The *continuous interval detector* (CID) operates on the same principle as the sampled interval detector, except that each hypothesis is modelled as a continuous-time process which is realised in a sampled domain. The performance of the sampled interval detector presented in the last section was shown to depend critically upon the band frequency and sample rate. Setting the band frequency too high or the sample rate too low, even to a modest extent, rendered the detector completely ineffective. The continuous interval detector constitutes an attempt to nullify this problem without resorting to wholesale up-sampling or other costly processing.

The motivation behind the continuous interval detector originates in consideration of the crucial role played by the sample rate in the detector performance. If the signal were up-sampled by a factor of two, then the resolution of the density functions would double and their intersection would be more accurately sampled. Pursuing this line of thought to its limit raises the question: Can the detection problem be formulated as though the signal were continuous? The next two sections answer this question in the affirmative; others have also successfully modelled the interval between the zeros of a continuous Gaussian process, notably Rice (1944).

The Continuous Interval Statistic

Having established that it is possible to reformulate the model in continuous time, we are left with the problem of having, in reality, to work with a sampled signal. However, there is no longer any requirement that the signal be up-sampled in its entirety; we only require accurate estimates of the zero crossing times. One possibility is to interpolate linearly between the samples of a zero crossing. For the two samples values $x_{j,0}$ and $x_{j,1}$ which make up the j-th crossing, the fractional crossing time is

$$Z_j = \frac{x_{j,0}}{x_{j,0} - x_{j,1}}, \tag{4.50}$$

from which a continuous estimate for the interval duration, I_c, is found:

$$I_c = (I - Z_0 + Z_1)\triangle t. \tag{4.51}$$

Here, $\triangle t$ denotes the sampling interval, in seconds. The relationships amongst these quantities is depicted in Figure 4.15. Assuming that H_0 and H_1 are equiprobable, the continuous interval detector operates according to the rule

$$\text{choose } H_1 \text{ iff } \frac{p_{I_c}(i_c \mid H_1)}{p_{I_c}(i_c \mid H_0)} > 1, \text{ otherwise choose } H_0. \tag{4.52}$$

The extraction of the test statistic prior to the decision rule is shown as a block diagram in Figure 4.16. Here, another specialised block has been introduced, labelled '`interp`', which corresponds to (4.50).

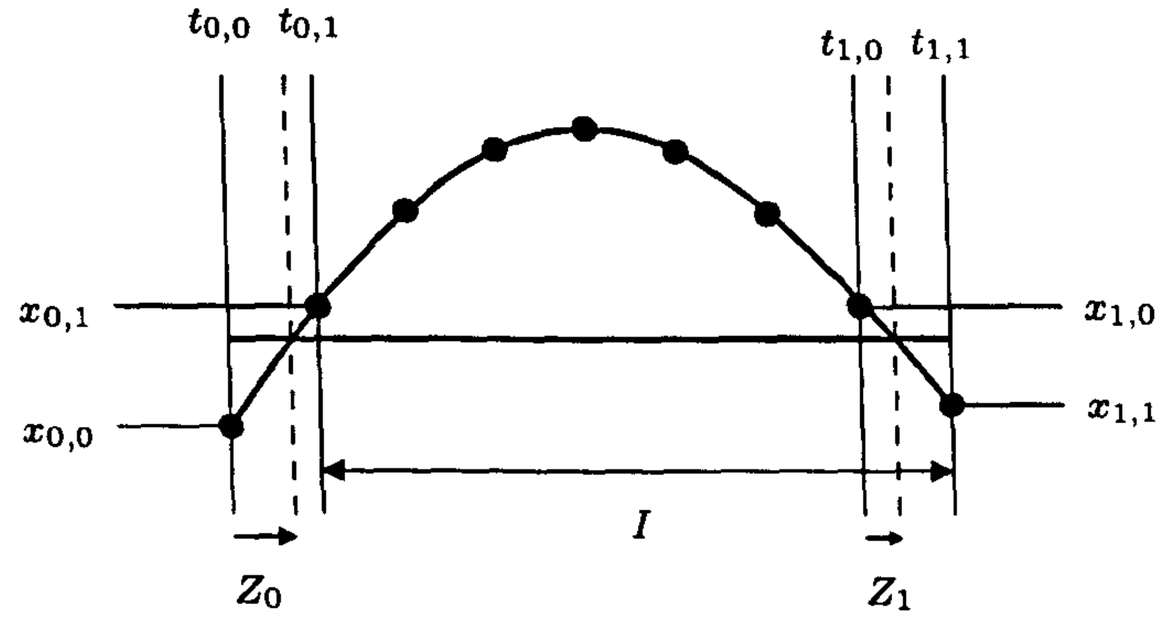


Figure 4.15: Linear interpolation of zero crossings.

4.4.2 Continuous-time Random Processes

The sampled interval detector dealt with sampled impulse responses, power spectral densities and autocorrelation functions. The aim of this section is to review how these quantities are translated into their continuous-time counterparts. In order to maintain a distinction between the two domains—sampled and continuous—the convention of surrounding discrete and continuous arguments with square and round parentheses was adopted, e.g., $h[n]$ and $h(t)$.

Two impulse responses characterise the simple systems considered so far: the impulse response of the analysis filter and the impulse response used to generate the signal process. The impulse response of the analysis filter is specified by $h_a(t)$, and the signal process is formed by convolving a white Gaussian noise signal with the impulse response $h_s(t)$, where, as before,

$$\int_{-\infty}^{\infty} h_s^2(t)dt = 1. \tag{4.53}$$

The continuous system defined by these impulse responses can be converted to a discrete-time system using an impulse invariant transform, a procedure which replaces the continuous impulse responses with sampled versions, i.e.,

$$h_a[n] = h_a(\Delta t \cdot n) \tag{4.54}$$

$$h_s[n] = h_s(\Delta t \cdot n). \tag{4.55}$$

The squared-magnitude response of the analysis filter, expressed in angular frequency, is related to the impulse response by

$$|\mathscr{H}_a(\omega)|^2 = \left| \int_{-\infty}^{\infty} h_a(t) e^{-i\omega t} dt \right|^2, \tag{4.56}$$

and that of the signal filter, $|\mathscr{H}_s(\omega)|^2$, is obtained from $h_s(t)$ in the same way. The power spectral density for the noise-only hypothesis, H_0, is the product of a white spectrum and $|\mathscr{H}_a(\omega)|^2$, that is,

$$\mathscr{S}_0(\omega) = \frac{N_0}{2} |\mathscr{H}_a(\omega)|^2, \tag{4.57}$$

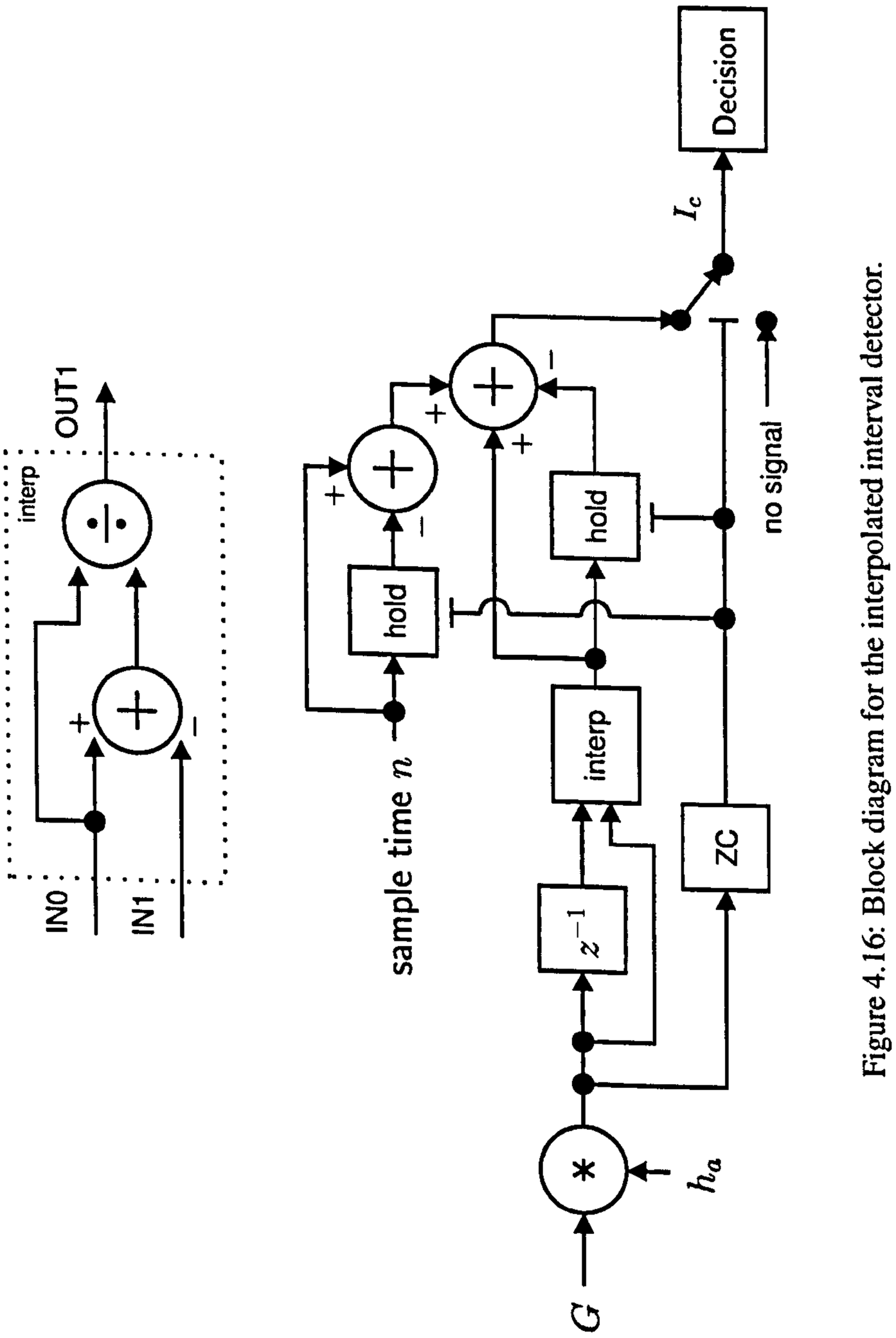


Figure 4.16: Block diagram for the interpolated interval detector.

where $N_0/2$ is a constant noise power spectral density. Meanwhile, the power spectral density for the signal-and-noise hypothesis, H_1, is

$$\mathscr{S}_1(\omega) = |\mathscr{H}_a(\omega)|^2 \left[\frac{N_0}{2} + \sigma_s^2 |\mathscr{H}_s(\omega)|^2\right], \tag{4.58}$$

where σ_s^2 is the total signal power, implying the signal-to-noise ratio

$$\text{SNR} = 10 \log_{10} \frac{\sigma_s^2}{N_0}, \text{ dB}. \tag{4.59}$$

As before, the autocovariance functions for hypothesis H_j is obtained via the inverse Fourier transform (an application of the Wiener-Khinchin relation)

$$\gamma_j(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathscr{S}_j(\omega) e^{i\omega t} d\omega, \tag{4.60}$$

and we arrive at the autocorrelation function $\rho_j(\tau)$ by dividing (4.60) by $\gamma_j(0)$. For the sampled interval detector, the conditional interval c.d.f. and p.d.f. were shown to rely solely upon the conditional autocorrelation, i.e., $p_I[i \mid H_j] \sim \rho_j[k]$. The next section takes the same step in the continuous domain: acquiring $p_{I_c}(i_c \mid H_j)$ from $\rho_j(\tau)$.

4.4.3 Probability Density Functions

In Section 4.3.2, the cumulative distribution function for sampled intervals was found to be

$$P(I \leq k) = \begin{cases} \frac{1}{2} + \frac{\sin^{-1}\rho[k+1] - \sin^{-1}\rho[k]}{\pi - 2\sin^{-1}\rho[1]} & k_0 < k < 2k_0 \\ 0 & k \leq k_0 \\ 1 & k \geq 2k_0. \end{cases} \tag{4.61}$$

Assuming a particular sampling interval Δt, by making use of rule $\rho[k] \approx \rho(\Delta t \cdot k)$, $\tau = \Delta t \cdot k$, the portion of (4.61) for which $k_0 < k < 2k_0$ can be rewritten

$$P(I \leq \tau) = \frac{1}{2} + \frac{\sin^{-1}\rho(\tau+\Delta t) - \sin^{-1}\rho(\tau)}{\pi - 2\sin^{-1}\rho(\Delta t)}, \ \tau/\Delta t \in \mathbb{N}. \tag{4.62}$$

The expression for the continuous c.d.f. emerges as the sampling interval tends to zero (i.e., $\Delta t \to 0$ and $f_s \to \infty$); we may therefore state the c.d.f. as the limit

$$P(I_c \leq \tau) = \lim_{\Delta t \to 0} \left\{ \frac{1}{2} + \frac{\sin^{-1}\rho(\tau+\Delta t) - \sin^{-1}\rho(\tau)}{\pi - 2\sin^{-1}\rho(\Delta t)} \right\} \tag{4.63}$$

$$= \frac{1}{2} + \lim_{\Delta t \to 0} \left\{ \frac{\sin^{-1}\rho(\tau+\Delta t) - \sin^{-1}\rho(\tau)}{\pi - 2\sin^{-1}\rho(\Delta t)} \right\}. \tag{4.64}$$

When solving the limit, it helps to multiply firstly the numerator and denominator by $\triangle t$ and then split it into a product of limits,

$$P(I_c \leq \tau) = \frac{1}{2} + \lim_{\triangle t \to 0} \left\{ \frac{\triangle t(\sin^{-1}\rho(\tau+\triangle t) - \sin^{-1}\rho(\tau))}{\triangle t(\pi - 2\sin^{-1}\rho(\triangle t))} \right\} \tag{4.65}$$

$$= \frac{1}{2} + \lim_{\triangle t \to 0} \left\{ \frac{\triangle t}{\pi - 2\sin^{-1}\rho(\triangle t)} \right\}$$

$$\times \lim_{\triangle t \to 0} \left\{ \frac{\sin^{-1}\rho(\tau+\triangle t) - \sin^{-1}\rho(\tau)}{\triangle t} \right\}. \tag{4.66}$$

Rice's Formula for the expected number of zero crossings in unit time (Rice, 1944) asserts the following as part of its solution:

$$\lim_{\triangle t \to 0} \left\{ \frac{\frac{1}{2} - \frac{1}{\pi}\sin^{-1}\rho(\triangle t)}{\triangle t} \right\} = \frac{1}{\pi}\sqrt{-\rho''(0)}. \tag{4.67}$$

The first limit in (4.66) can be reworked using (4.67) to give

$$P(I_c \leq \tau) = \frac{1}{2} + \frac{1}{2\sqrt{-\rho''(0)}} \lim_{\triangle t \to 0} \left\{ \frac{\sin^{-1}\rho(\tau+\triangle t) - \sin^{-1}\rho(\tau)}{\triangle t} \right\}. \tag{4.68}$$

The second limit expresses the derivative of the arcsine of the autocorrelation function and may be solved directly[1], leaving the continuous cumulative distribution function in terms of the autocorrelation function $\rho(\tau)$ and its first two derivatives:

$$P(I_c \leq \tau) = \begin{cases} \frac{1}{2} + \frac{\rho'(\tau)}{2\sqrt{\rho''(0)(\rho^2(\tau)-1)}} & \tau_0 < k < 2\tau_0 \\ 0 & \tau \leq \tau_0 \\ 1 & \tau \geq 2\tau_0. \end{cases} \tag{4.69}$$

The continuous interval probability density function is found by differentiating (4.69), with respect to interval duration i_c, applying the quotient rule[2],

$$p_{I_c}(i_c) = \begin{cases} \frac{(\rho^2(i_c)-1)\rho''(i_c) - \rho(i_c)(\rho'(i_c))^2}{2(\rho^2(i_c)-1)^{3/2}\sqrt{\rho''(0)}} & \tau_0 < i_c < 2\tau_0 \\ 0 & \text{otherwise.} \end{cases} \tag{4.70}$$

Comparing the Analytical and Empirical Distributions

The previous two sections outlined the stages required to transform the impulse responses of the system into interval probability density functions. In this section, a controlled random signal is generated and a histogram is formed from continuous

[1] using $\frac{d}{dx}\{\sin^{-1}x\} = \frac{1}{1-x^2}$.

[2] for numerical stability, compute the denominator using $2(\rho^2(i_c)-1)\sqrt{\rho''(0)(\rho^2(i_c)-1)}$.

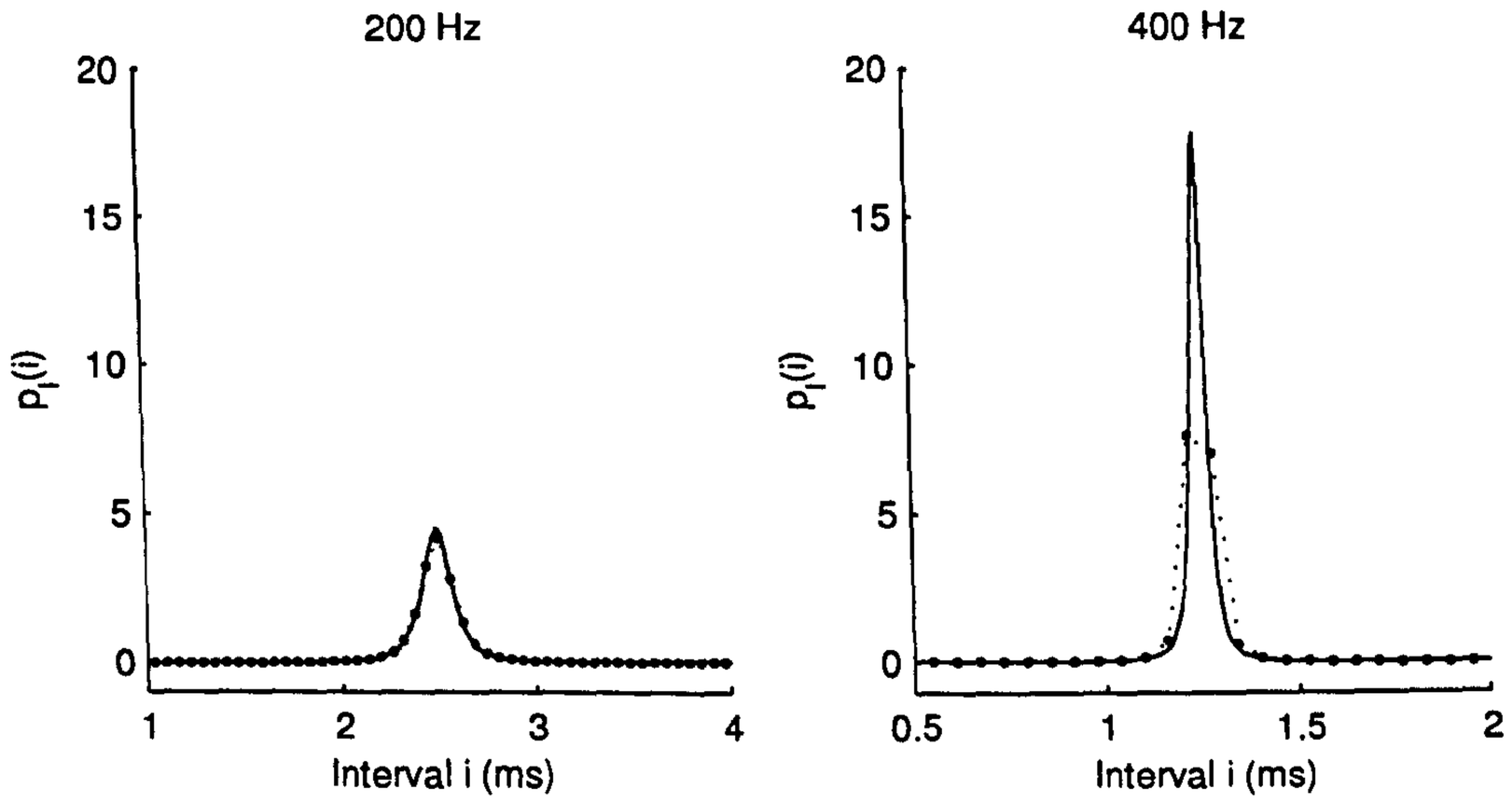


Figure 4.17: Analytical and empirical probability density for continuous and sampled intervals: continuous (solid, dash-dotted lines) and sampled (dotted line, solid circles).

intervals estimated by linear interpolation. The histogram is then compared with the continuous interval density function predicted by (4.70), as was done for the sampled interval detector in Section 4.3.2. The purpose of this comparison is to determine whether the analytical probability density function agrees with observed data, and, if so, whether it offers high resolution at high frequencies.

The continuous model filter used to generate the random process in question has the impulse response

$$h_a(t) = \exp(-2(40t)^2)\cos(2\pi f_a t) \tag{4.71}$$

in which f_a is set to either 200 Hz or 400 Hz. For the sample rate $f_s = 16384$, an impulse-invariant transform yields a sampled (and truncated) version of the impulse response

$$h_a[n] = \begin{cases} \exp(-2(40n/f_s)^2)\cos(2\pi f_a n/f_s) & -1024 \leq n \leq 1024 \\ 0 & \text{otherwise.} \end{cases} \tag{4.72}$$

The discrete-time random process used to produce the histogram is generated by convolving white Gaussian noise with $h_a[n]$. The continuous model autocorrelation function is very closely approximated by

$$\rho(\tau) = \exp(-(40\tau)^2)\cos(2\pi f_a \tau). \tag{4.73}$$

The continuous interval p.d.f. is found by placing $\rho(\tau)$ above into (4.70). The analytical and empirical probability density functions for the continuous interval statistic i_c are plotted in Figure 4.17 for band frequencies 200 Hz and 400 Hz at a resolution, i.e, histogram bin width, of 5 μs. The analytical sampled interval p.d.f.s, obtained from (4.35), are rescaled and overlaid for comparison with a resolution of $\triangle t \approx 62$ μs.

4.4.4 Modulated Gaussian Mixture Models (MGMMs)

The interval cumulative distribution and probability density functions for a hypothesis H_j are defined exclusively in terms of the autocorrelation function $\rho_j(\tau)$ and its first two derivatives, $\rho'_j(\tau)$ and $\rho''_j(\tau)$. In order to determine the autocorrelation function of the entire system in response to white Gaussian noise (or, indeed, any wide-sense stationary random processes), the impulse responses of all its subsystems must undergo several transformations. A detailed treatment of signals and systems has been provided by many other authors and is omitted here.

In principle, the autocorrelation function of the signal at any point in a linear system can always be determined analytically for jointly wide-sense stationary inputs. In practice, finding an expression for the autocorrelation depends on the availability of closed-form expressions for the Fourier transform or convolution integrals that necessarily arise. The following discussion offers a parametric format for a continuous linear system, which guarantees that an exact, closed-form expression exists for the autocorrelation, interval cumulative distribution and probability density functions.

The basic building-block of the model described next is the modulated Gaussian component (MGC). A MGC Ψ consists of five parameters

$$\Psi = \langle A, C, \mu, \bar{\omega}, \phi \rangle$$

where $C \geq 0$, and is realised as a complex signal by

$$\Psi \models x(t) = A \exp(-2(C(t-\mu))^2) \cdot \exp(i\bar{\omega}t + i\phi). \tag{4.74}$$

This is readily recognised as a complex modulated Gaussian pulse, encountered widely in the sonar literature (Burdic, 1984). All the parameters have a physical interpretation: A controls amplitude; C controls the width of the pulse; μ specifies a shift in time; $\bar{\omega}$ is the radial frequency; ϕ is a phase shift. The *models* symbol ($\models$) is adopted here as a convenient means of switching, in either direction, between the description of a model and its realisation as a signal. The identity of two MGCs is established on the basis of whether they realise the same signal, as opposed to whether their parameters are the same. The set of all MGCs is denoted $\mathbf{\Psi}$.

A modulated Gaussian mixture model (MGMM) is a sum of modulated Gaussian components and is described by the set of its components

$$\Lambda = \{\Psi_1, \Psi_2, \ldots, \Psi_N\},\ \forall j \in 1 \ldots N, \Psi_j \in \mathbf{\Psi}$$

and is realised by the sum:

$$\Lambda \models x(t) = \sum_{j=1}^{N} A_j \exp(-2(C_j(t-\mu_j))^2) \cdot \exp(i\bar{\omega}_j t + i\phi_j) \models \Lambda, \tag{4.75}$$

where the subscript j references the parameter of the j-th component. It should be noted that the term "modulated Gaussian mixture model" has a separate interpretation to the "Gaussian mixture model" (GMM) frequently employed as a model distribution. A Gaussian mixture model is a parametric description of a probability density function

(Bishop and Hinton, 1995), whereas an MGMM is intended to model a signal rather than a density function. That the terminology should resurface in another context is largely due to the convenient properties of the Gaussian function itself. We shall denote the set of all possible MGMMs using $\mathbf{\Lambda}$. Two MGMMs are considered identical if they realise identical signals. (It is tentatively suggested that a "canonical" form exists for any MGMM, in which every component satisfies $A > 0$, and no two components have identical $\bar{\omega}$.)

The following sections define a series of operators for MGCs and MGMMs. These operators have been implemented in MATLAB as functions which perform operations on data structure arrays with the fields `A`, `C`, `mu`, `w` and `phi`. The advantage here is extensibility. For instance, deriving the autocorrelation function for a linear system comprising many subsystems is laborious, owing in particular to the large number of cross-terms that emerge during the working. Constraining the impulse response of each system to a MGMM permits functions such as the autocorrelation to be computed in "short-hand" and in a mechanical fashion.

Scaling

The scaling of a MGC Ψ_1 by a real constant α produces another MGC Ψ_2.

$$\begin{aligned} \alpha\Psi_1 \in \mathbf{\Psi} &= \alpha \langle A_1, C_1, \mu_1, \bar{\omega}_1, \phi_1 \rangle && (4.76)\\ &\models \alpha A_1 \exp(-2(C_1(t-\mu_1))^2) \cdot \exp(i\bar{\omega}_1 t + i\phi_1) && (4.77)\\ &\models \langle A_2 = \alpha A_1, C_1, \mu_1, \bar{\omega}_1, \phi_1 \rangle \in \mathbf{\Psi} && (4.78)\\ &= \Psi_2 \in \mathbf{\Psi}. && (4.79) \end{aligned}$$

Multiplying a MGMM by a real constant scales each of its components, so MGMMs are closed on the operation of scalar multiplication. It is easy to show that both MGCs and MGMMs are also closed on multiplication by a complex constant.

Addition and Subtraction

The addition of two MGMMs Λ_1 and Λ_2 produces a third MGMM. The rule used to add them together is defined recursively:

$$\begin{aligned} \Lambda_1 + \emptyset &= \Lambda_1 \\ \Lambda_1 + \Lambda_2 &= (\Lambda_1 \oplus \Lambda_2) + \{\langle 2A, C, \mu, \bar{\omega}, \phi \rangle : \langle A, C, \mu, \bar{\omega}, \phi \rangle \in \Lambda_1 \cap \Lambda_2\}. \end{aligned}$$

where $\oplus$ denotes mutual set difference. Simply taking a union of components would imply that duplicates appear only once in the resulting MGMM. This definition ensures that duplicate elements are *added*. A simple example can illustrate this point:

$$\begin{aligned} &\{\langle 4,10,0,200,0\rangle, \langle 2,10,0,200,0\rangle\} + \{\langle 3,12,0,100,6\rangle, \langle 2,10,0,200,0\rangle\} \\ &\quad = \{\langle 4,10,0,200,0\rangle, \langle 3,12,0,100,6\rangle\} + \{\langle 4,10,0,200,0\rangle\} \\ &\quad = \{\langle 3,12,0,100,6\rangle\} + \{\langle 8,10,0,200,0\rangle\} \\ &\quad = \{\langle 3,12,0,100,6\rangle, \langle 8,10,0,200,0\rangle\} + \emptyset \\ &\quad = \{\langle 3,12,0,100,6\rangle, \langle 8,10,0,200,0\rangle\}. \end{aligned}$$

Because it is possible to multiply a MGMM by -1, MGMMs are closed on subtraction as well as addition.

Multiplication

It is slightly more difficult to demonstrate that the product of two MGCs is a MGC. The best way is to split the product up into its scalar, Gaussian and phasor factors:

$$\begin{aligned}\Psi_1 \cdot \Psi_2 \quad \models \quad & [A_1 A_2] \left[\exp(-2(C_1(t-\mu_1))^2) \exp(-2(C_2(t-\mu_2))^2)\right] \\ & \times \left[\exp(i\bar{\omega}_1 t + i\phi_1) \exp(i\bar{\omega}_2 t + i\phi_2)\right] \qquad (4.80) \\ = \quad & A_1 A_2 \left[\exp(-2(C_1(t-\mu_1))^2) \exp(-2(C_2(t-\mu_2))^2)\right] \\ & \times \exp(i(\bar{\omega}_1 + \bar{\omega}_2)t + i(\phi_1 + \phi_2)). \qquad (4.81)\end{aligned}$$

The scalar factors multiply to give a scalar; the phasor factors multiply to give a phasor; what of the Gaussian factor? This shall be tackled separately:

$$\begin{aligned}& \exp(-2(C_1(t-\mu_1))^2) \exp(-2(C_2(t-\mu_2))^2) \qquad (4.82) \\ = \quad & \exp\left(-2\left[C_1^2(t-\mu_1)^2 + C_2^2(t-\mu_2)^2\right]\right) \qquad (4.83) \\ = \quad & \exp\left(-2\left[C_1^2(t^2+\mu_1^2-2\mu_1 t) + C_2^2(t^2+\mu_2^2-2\mu_2 t)\right]\right) \qquad (4.84) \\ = \quad & \exp\left(-2\left[(C_1^2+C_2^2)t^2 - 2(C_1^2\mu_1 + C_2^2\mu_2)t + C_1^2\mu_1^2 + C_2^2\mu_2^2\right]\right) \qquad (4.85)\end{aligned}$$

Using the abbreviation $D = C_1^2 + C_2^2$ and completing the square in the exponential term gives

$$\begin{aligned}& \exp\left(-2D\left[t^2 - 2\frac{(C_1^2\mu_1 + C_2^2\mu_2)}{D}t\right] - 2\left[C_1^2\mu_1^2 + C_2^2\mu_2^2\right]\right) \qquad (4.86) \\ = \quad & \exp\left(-2D\left[t - \frac{(C_1^2\mu_1 + C_2^2\mu_2)}{D}\right]^2\right) \\ & \times \exp\left(\frac{2(C_1^2\mu_1 + C_2^2\mu_2)^2}{D} - 2\left[C_1^2\mu_1^2 + C_2^2\mu_2^2\right]\right). \qquad (4.87)\end{aligned}$$

We see that (4.87) has Gaussian form. Therefore, the product of two MGCs Ψ_1 and Ψ_2 is another MGC, $\Psi_3 =$

$$\begin{aligned}\langle A_3 \quad = \quad & A_1 A_2 \exp\left(\frac{2(C_1^2\mu_1 + C_2^2\mu_2)^2}{C_1^2 + C_2^2} - 2\left[C_1^2\mu_1^2 + C_2^2\mu_2^2\right]\right), \\ C_3 \quad = \quad & \sqrt{C_1^2 + C_2^2}, \\ \mu_3 \quad = \quad & \frac{C_1^2\mu_1 + C_2^2\mu_2}{C_1^2 + C_2^2}, \\ \bar{\omega}_3 \quad = \quad & \bar{\omega}_1 + \bar{\omega}_2, \\ \phi_3 \quad = \quad & \phi_1 + \phi_2 \rangle .\end{aligned}$$

One can immediately infer from this result, that the product of two MGMMs is a MGMM. Any MGMM equals the sum of its individual components, as defined by

MGMM addition above. Hence, expanding the brackets,

$$\begin{aligned}\Lambda_1 \cdot \Lambda_2 &= (\Psi_{1,1} + \Psi_{1,2} + \ldots + \Psi_{1,N}) \cdot (\Psi_{2,1} + \Psi_{2,2} + \ldots + \Psi_{2,M}) \\ &= \Psi_{1,1} \cdot \Psi_{2,1} + \Psi_{1,2} \cdot \Psi_{2,1} + \ldots + \Psi_{1,N} \cdot \Psi_{1,M} \quad (4.88) \\ &= \Lambda_3. \quad (4.89)\end{aligned}$$

The product of two MGMMs with M and N components may contain as many as MN components, although a canonical version may contain fewer. Note that the multiplicative identity element for multiplication is $\langle A{=}1, C{=}0, \mu{=}0, \bar{\omega}{=}0, \phi{=}0\rangle$ and that division is not defined for MGMMs in general.

Real Part, Imaginary Part and Squared-Magnitude

By applying Euler's formula to each component, the real part of a MGMM can be shown to be a MGMM:

$$\begin{aligned}\mathrm{Re}\{\Lambda_1\} &\models \mathrm{Re}\left\{\sum_{j=1}^{N} A_j \exp(-2(C_j(t-\mu_j))^2)\exp(i\bar{\omega}_j t + i\phi_j)\right\} \quad (4.90) \\ &= \sum_{j=1}^{N} A_j \exp(-2(C_j(t-\mu_j))^2)\cos(\bar{\omega}_j t + \phi_j) \quad (4.91) \\ &= \sum_{j=1}^{N} \frac{A_j}{2} \exp(-2(C_j(t-\mu_j))^2)[e^{i\bar{\omega}_j t + i\phi_j} + e^{-i\bar{\omega}_j t - i\phi_j}] \quad (4.92) \\ &\models \Lambda_2. \quad (4.93)\end{aligned}$$

The imaginary part, $\mathrm{Im}\{\Lambda_1\}$, is also a MGMM. As both the real and imaginary parts are both MGMMs, it follows that the squared magnitude is also a MGMM,

$$|\Lambda_2|^2 = \mathrm{Re}\{\Lambda_2\}^2 + \mathrm{Im}\{\Lambda_2\}^2 = \Lambda_3. \quad (4.94)$$

Fourier Transform

The Fourier transform is linear so we need only consider the effect of the transform upon an individual MGC. The Fourier transform of a zero-mean Gaussian is another zero-mean Gaussian,

$$\begin{aligned}\mathcal{F}\{\exp(-2(Ct)^2)\} &= \int_{\infty}^{\infty} \exp(-2(Ct)^2)e^{-i\omega t}dt \quad (4.95) \\ &= \int_{\infty}^{\infty} \exp\left(-2C^2\left[t^2 + \frac{i\omega t}{C^2}\right]\right)dt \quad (4.96) \\ &= \int_{\infty}^{\infty} \exp\left(-2C^2\left[t + \frac{i\omega}{2C^2}\right]^2 - \frac{\omega^2}{2C^2}\right)dt \quad (4.97) \\ &= \frac{\sqrt{2\pi}}{2C}\exp\left(-\frac{\omega^2}{2C^2}\right). \quad (4.98)\end{aligned}$$

Using this result, in conjunction with the Fourier time and frequency shift theorems, it is a straight-forward matter to show that

$$\mathcal{F}\{\Lambda_1(t)\} = \left\langle \begin{array}{rcl} A_2 & = & A_1\sqrt{2\pi}/(2C_1), \\ C_2 & = & 1/(4C_1), \\ \mu_2 & = & \bar{\omega}_1, \\ \bar{\omega}_2 & = & -\mu_1, \\ \phi_2 & = & \phi_1 \end{array} \right\rangle = \Lambda_2(\omega). \tag{4.99}$$

Hence, the Fourier transform of a MGMM is another MGMM with the same number of components, expressed in the frequency domain (in radial units ω). Note that in order for the Fourier transform to be defined for components with $C = 0$, we must allow that $C = +\infty$, which in effect models a Dirac delta.

Further Operations

It has been shown that MGMMs are closed on multiplication by a scalar, addition (also subtraction and negation), multiplication, the real, imaginary and squared-magnitude operations, and the Fourier transform. By combining these, it is immediately evident that MGMMs are also closed on the following operations: raising to a positive integer power (repeated application of multiplication); the inverse Fourier transform (apply the duality property of the Fourier transform); and convolution (forward and inverse Fourier transform; multiplication).

In addition, the form of the MGMM itself caters for standard transformations such as reversal, shifting, dilation and complex conjugation in the time or frequency domain. Also, integrating any MGMM on the bounds $[-\infty, \infty]$ is a simple matter. Thus we arrive at our goal: if the impulse responses of continuous sub-systems within a larger linear system are specified as MGMMs, then combinations of the operations described above are guaranteed to find closed-form expressions for the system impulse response, autocovariance and autocorrelation functions, power spectral densities, complex and squared-magnitude responses, and ultimately, the continuous interval p.d.f..

4.4.5 Setting up the Experiments

The narrowband detection experiments described in Section 4.1 can now be performed for the continuous interval detector under the same conditions as the squared-envelope and sampled interval detectors. Formerly, obtaining the autocorrelation function for the signal-and-noise hypothesis would have been quite tedious; now, we can specify the impulse responses as MGMMs and utilise the MGMM operations to transform the model as required.

Analysis Filter, Signal Process and Hypotheses

The model impulse response used to generate the signal is

$$h_s(t) = \exp\left[-2\left(\frac{\alpha_s}{T_s}t\right)^2\right]\cos(2\pi \cdot f_c t) \tag{4.100}$$

$$\models \left\langle \frac{1}{2}, \frac{\alpha_s}{T_s}, 2\pi f_c, 0, 0 \right\rangle + \left\langle \frac{1}{2}, \frac{\alpha_s}{T_s}, -2\pi f_c, 0, 0 \right\rangle \tag{4.101}$$

$$= \Lambda_{h_s}. \tag{4.102}$$

which is then normalised to pass unit power, as before,

$$\Lambda_{h_s} \leftarrow \frac{\Lambda_{h_s}}{\sqrt{\int |\Lambda_{h_s}|^2 dt}}. \tag{4.103}$$

The analysis filter impulse response has an identical form, but is parameterised by α_a, T_a and f_a. The power spectral density models for H_0 and H_1 are then found by

$$\Lambda_{\mathscr{S}_0} = |\mathcal{F}\{\Lambda_{h_a}\}|^2 \frac{N_0}{2} \tag{4.104}$$

$$\Lambda_{\mathscr{S}_1} = |\mathcal{F}\{\Lambda_{h_a}\}|^2 \left(\frac{N_0}{2} + \sigma_s^2 |\mathcal{F}\{\Lambda_{h_s}\}|^2\right), \tag{4.105}$$

as are the model autocovariance and autocorrelation functions,

$$\Lambda_{\gamma_j} = \mathcal{F}^{-1}\{\Lambda_{S_j}\} \tag{4.106}$$

$$\Lambda_{\rho_j} = \frac{\Lambda_{\gamma_j}}{\Lambda_{\gamma_j}(0)}. \tag{4.107}$$

Constructing the Continuous Interval Detector

Assuming that H_0 and H_1 are equiprobable, the continuous interval detector operates according to the rule

$$\text{choose } H_1 \text{ iff } \frac{p_{I_c}(i_c \mid H_1)}{p_{I_c}(i_c \mid H_0)} > 1, \text{ otherwise choose } H_0,$$

where i_c is a continuous interval test statistic computed from linear-interpolated zero crossings and $p_{I_c}(i_c|H_j)$ are conditional probability density functions obtained by evaluating Λ_{ρ_j} in (4.70).

The probability of error can be calculated analytically by finding the decision regions corresponding to the minimum error criterion and then integrating the p.d.f. in these regions using the continuous cumulative distributions function (4.70). Computing the minimum error decision boundaries is difficult however, as it requires the intersections of the conditional p.d.f.s to be located. A better strategy is to closely approximate the analytical probability of error by numerical integration. The analytical predictions presented next are obtained by applying (4.45) and (4.46) to the continuous interval p.d.f.s and using a very fine integration step, namely, 0.1 μs.

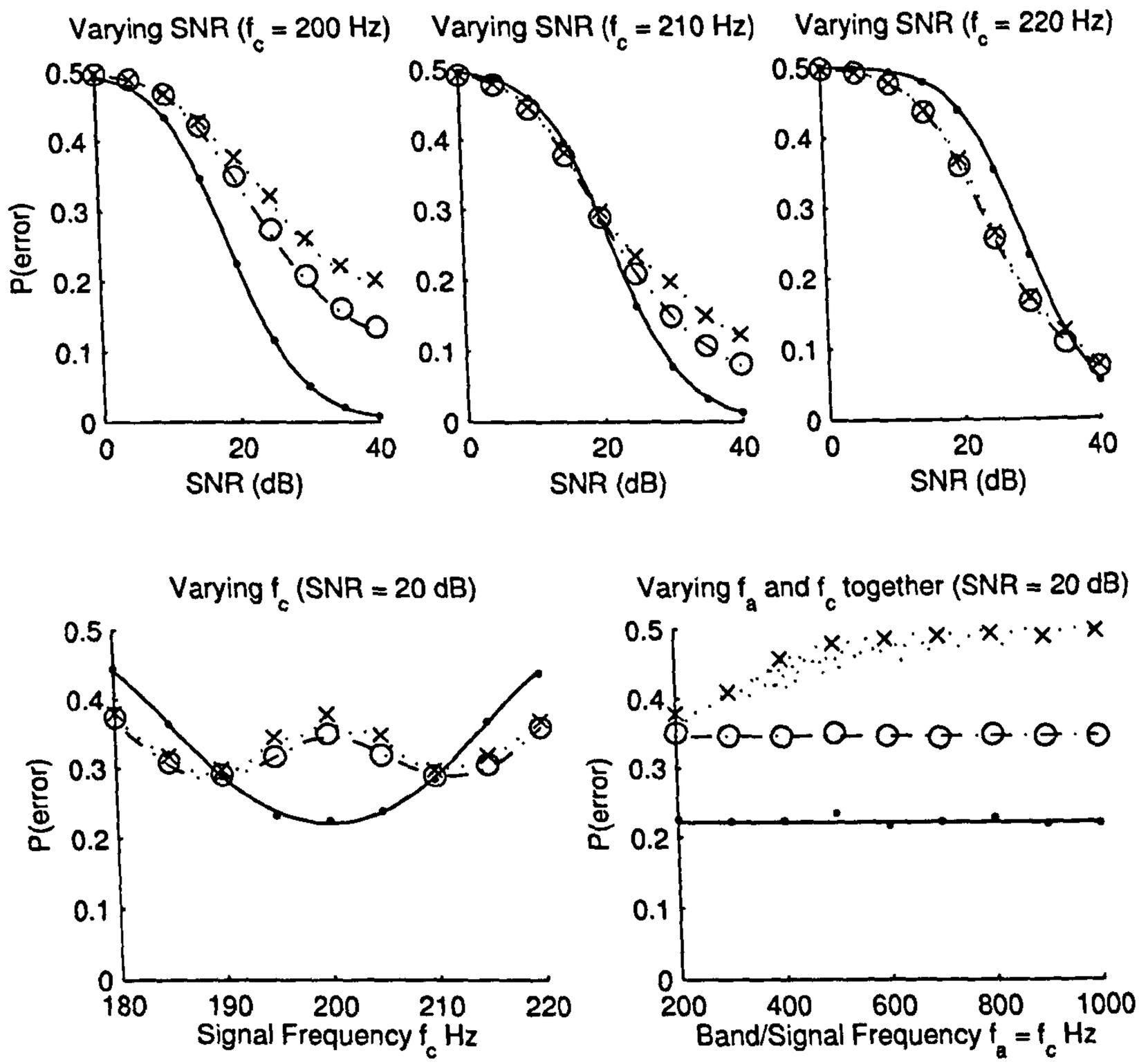


Figure 4.18: Probability of error in the detectors presented so far. The predicted and observed values are shown using lines and markers, according to the following key: squared-envelope detector (solid line; solid circle •); sampled interval detector (dotted line; cross ×) and continuous interval detector (dash-dotted line; open circle ○).

4.4.6 Experimental Results and Analysis

The results of the experiment above are provided in Figure 4.18, alongside those of the squared-envelope and sampled interval detectors.

How does the detector's performance vary with SNR? For the experiments in which the SNR is varied—the top three plots—the error probability curves for the continuous interval detector resemble, in terms of overall shape, those of the sampled interval detector. The continuous interval detector outperforms the sampled interval detector in all conditions, but the difference in performance is most pronounced when the signal is centred on the band, especially at high SNRs.

What is the effect of displacing the signal from the band centre? The bottom left-hand plot of Figure 4.18 shows that, as the signal frequency is removed from the band centre,

the error probability initially drops and then rises again with larger displacements. The trend in these results was explained in Section 4.3.5 for the sampled interval detector, and the same explanation applies to the continuous interval detector. Shifting the signal away from the centre of the analysis band causes the means of the conditional interval distributions to move apart, which initially reduces the probability of error. At the same time, however, the signal is attenuated further from the band centre and this leads to an increase in probability of error.

Does the detection performance relate to the absolute frequency of the signal? The unacceptably low performance of the sampled interval detector at high frequencies was discussed in Section 4.3.5 and shown there to result from the poor resolution of the interval probability density functions. The continuous interval detector was designed in response to this deficiency and, by assuming continuous models of the underlying random processes, replaces the sampled interval p.d.f.s with continuous versions. The results shown in the bottom right-hand graph of Figure 4.18 suggest that the solution works: the probability of error in the continuous interval detector is constant over the range 200–1000 Hz, unlike its sampled counterpart.

Is it possible to predict the performance of the detector? Yes. From a visual inspection of Figure 4.18, the empirical probability of error for the continuous interval detector appears to be correctly predicted by the numerical integration of the probability density functions described in Section 4.4.5 above.

How do the results of the different detectors compare? The squared-envelope detector still outperforms both interval detectors when the target signal is placed near the centre of the analysis band. The continuous interval detector improves upon the sampled interval detector in that its performance is unaffected by absolute shifts in frequency. It is reasonable to conjecture that the continuous interval detector represents the best single-interval, single-filter, timing-only detector that it is possible to construct, notwithstanding small errors incurred by the linear interpolation of zero crossing times. In light of this final remark, we turn our attention to the interpolated interval detector, which models the interpolation of zero crossings explicitly.

4.5 Interpolated Interval Detector

The sampled interval detector was shown to perform poorly at high frequencies, owing to rounding errors introduced by sampling intervals. The continuous interval detector offered an improved solution, generating continuous interval estimates by interpolating zero crossings and modelling the signal as a continuous-time process. This section introduces the *interpolated interval detector* (IID) as a hybrid: the test statistic is derived from linear interpolations; the likelihood test, however, explicitly models the conditional probability of linear-interpolated zero crossings in a sampled domain. Whether modelling the interpolations explicitly will lead to a detector that outperforms the continuous interval detector is unclear.

The interpolated interval detector is supplied with three measurements: the sampled zero crossing interval i, and the fractional zero crossing time of each zero crossing given by a linear interpolation, z_0 and z_1. The likelihood test has the following form

$$\text{choose } H_1 \text{ iff } \frac{p_{IZ_0Z_1}(i, z_0, z_1 \mid H_1)}{p_{IZ_0Z_1}(i, z_0, z_1 \mid H_0)} > 1, \text{ otherwise choose } H_0. \tag{4.108}$$

This decision rule is recognisable as that of the sampled interval detector, augmented with the fractional crossing times, z_j, to provide the information otherwise lost through rounding error. These quantities are indicated on the diagram in Figure 4.15. The task of building a minimum error detector amounts to determining the joint probability density for an observation $\langle i, z_0, z_1 \rangle$. The sampled interval and fractional crossing times are extracted in the same manner as the continuous interval detector, except that, instead of subtracting and adding the fractional crossing times to form the continuous interval statistic, the three components are fed separately to the decision rule, as the block diagram in Figure 4.19 shows.

4.5.1 Interpolated Crossing Probability

Before considering intervals, we examine the simpler problem of determining the probability (density) associated with a linear interpolation placing a zero crossing at fractional sample time Z. Two consecutive samples, x_0 and x_1, with respective sample times t_0 and t_1 are taken, and it is assumed that they are jointly Gaussian, with unit variance and correlation coefficient $\rho[1]$. Initially, we relax the constraint that the samples must form a crossing and permit extrapolations (see below). A linear interpolation of the two samples gives the fractional zero crossing time

$$Z = \frac{x_0}{x_0 - x_1}. \tag{4.109}$$

A zero crossing only occurs when x_0 and x_1 are unequal. If the product of x_0 and x_1 is negative or zero, then $0 \leq Z \leq 1$, and the zero crossing is *interpolated* (Figure 4.20A); otherwise the zero crossing is *extrapolated* (Figure 4.20B). The cumulative distribution

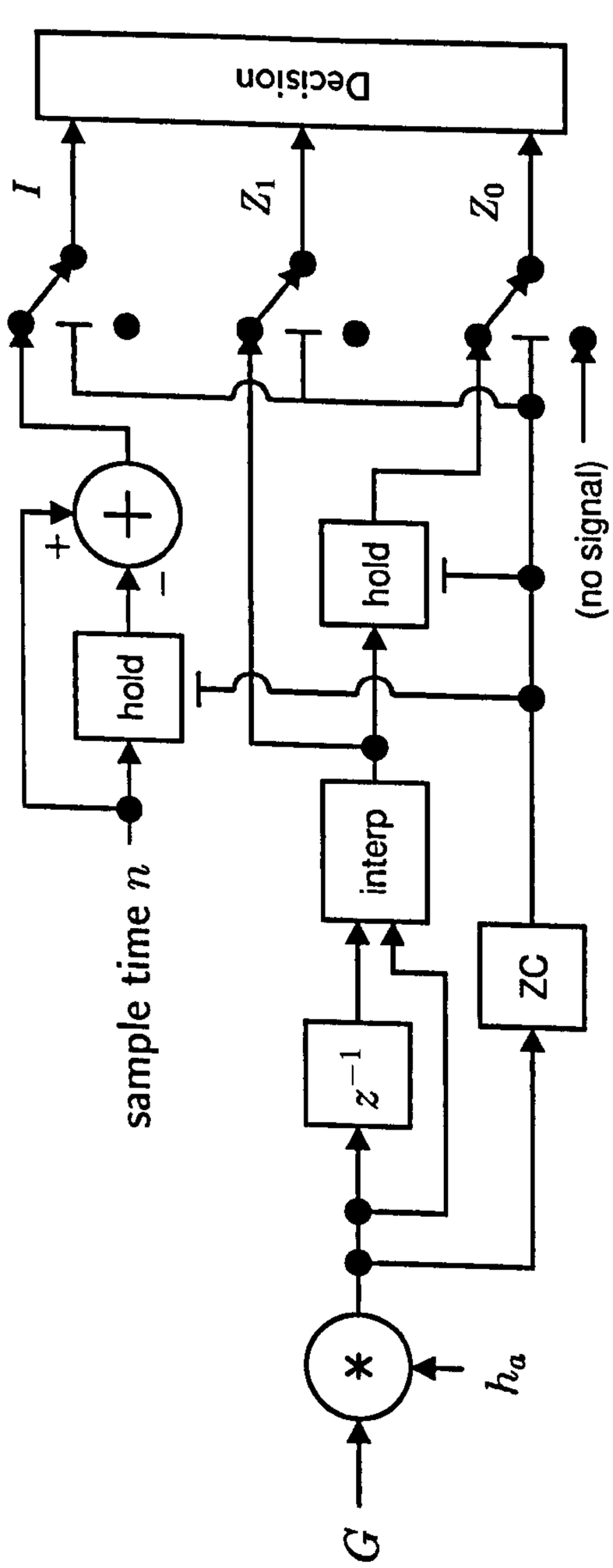


Figure 4.19: A block diagram for the interpolated interval detector.

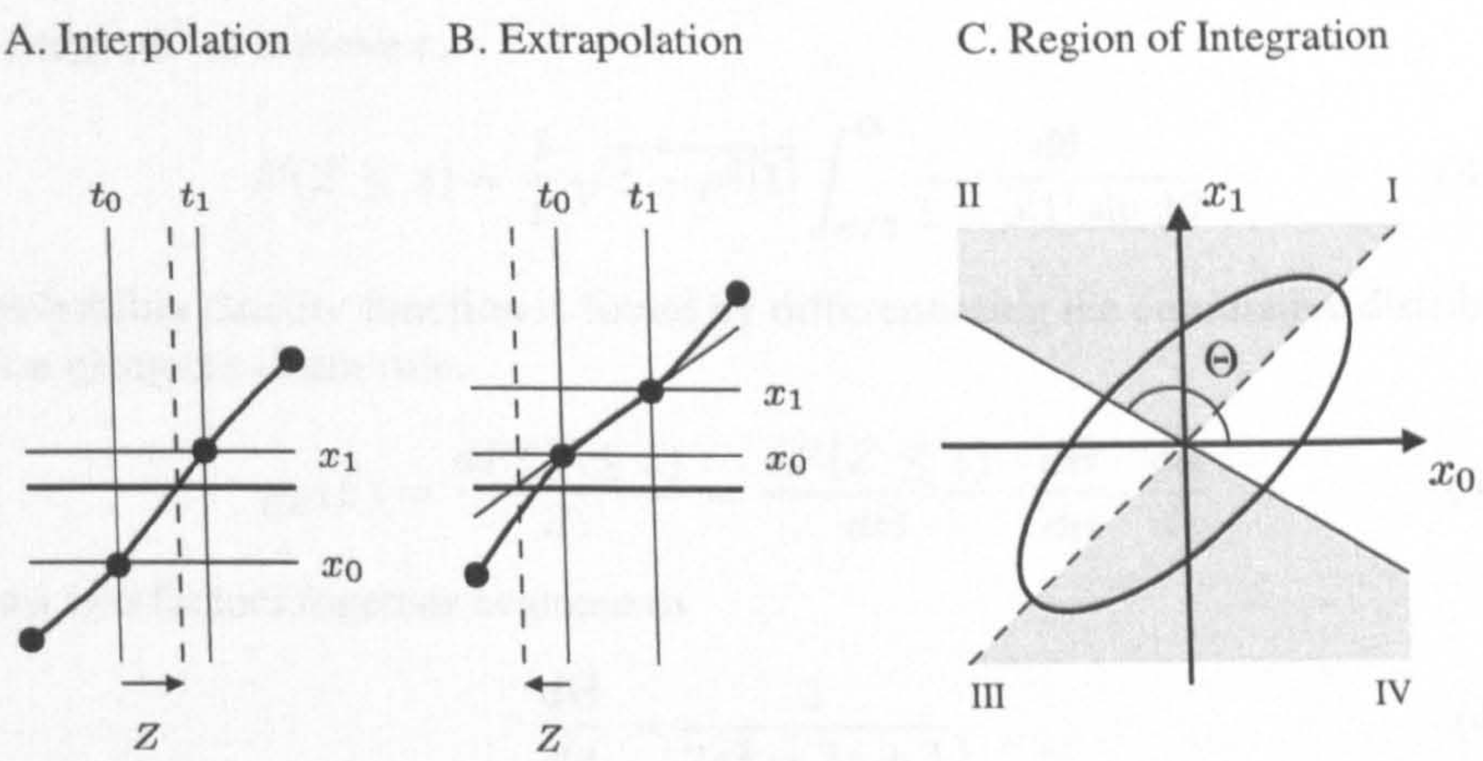


Figure 4.20: A) an interpolation; B) an extrapolation (the zero crossing is placed prior to t_0); C) the grey region indicates the event that corresponds to $Z \leq [1 - \tan\Theta]^{-1}$. The ellipse represents a contour of the p.d.f., the slope of the solid line is α, and the dashed line corresponds to $x_1 = x_0$.

function for Z is given by

$$P(Z \leq z) = P\left(\frac{x_0}{x_0 - x_1} \leq z\right) \tag{4.110}$$

$$= \begin{cases} P\left(x_1 \leq \alpha x_0, x_0 > x_1\right) & \\ + P\left(x_1 \geq \alpha x_0, x_0 < x_1\right) & z > 0 \\ P\left(x_1 \leq 0, x_0 > x_1\right) + P\left(x_1 \geq 0, x_0 < x_1\right) & z = 0 \\ P\left(x_1 \geq \alpha x_0, x_0 > x_1\right) & \\ + P\left(x_1 \leq \alpha x_0, x_0 < x_1\right) & z < 0, \end{cases}$$

where $\alpha = 1 - 1/z$, and represents the gradient of a straight line in x_0x_1-space which passes through the origin. The probability in (4.110) is evaluated by integrating the bivariate probability density function for $p_{x_0x_1}(\cdot)$ in the shaded region shown in Figure 4.20C, following a change to polar co-ordinates:

$$P(Z \leq z) = \tag{4.111}$$

$$\frac{1}{2\pi\sqrt{1-\rho^2[1]}}\left(\int_{\pi/4}^{\Theta} + \int_{\pi/4+\pi}^{\Theta+\pi}\right)\int_0^{\infty} r \exp\left(\frac{r^2(1-\rho[1]\sin 2\theta)}{-2(1-\rho^2[1])}\right) dr d\theta,$$

where $\Theta = \arctan\alpha$. Note that, due to symmetry, this simplifies to

$$P(Z \leq z) = \frac{1}{\pi\sqrt{1-\rho^2[1]}}\int_{\pi/4}^{\Theta}\int_0^{\infty} r \exp\left(\frac{r^2(1-\rho[1]\sin 2\theta)}{-2(1-\rho^2[1])}\right) dr d\theta. \tag{4.112}$$

Next, integrate[1] to remove r,

$$P(Z \leq z) = \frac{1}{\pi}\sqrt{1-\rho^2[1]}\int_{\pi/4}^{\Theta}\frac{d\theta}{1-\rho[1]\sin 2\theta}. \tag{4.113}$$

The probability density function is found by differentiating the cumulative distribution function using the chain rule.

$$p_Z(z) = \frac{dP(Z \leq z)}{dz} = \frac{dP(Z \leq z)}{d\Theta}\cdot\frac{d\Theta}{d\alpha}\cdot\frac{d\alpha}{dz}. \tag{4.114}$$

The last two factors together evaluate to

$$\frac{d\Theta}{dz} = \frac{1}{2z^2-2z+1}, \tag{4.115}$$

whilst the first factor—the derivative of the c.d.f. with respect to angle—is

$$\frac{1}{\pi}\sqrt{1-\rho^2[1]}\frac{d}{d\Theta}\left\{\int_{\pi/4}^{\Theta}\frac{d\theta}{1-\rho[1]\sin 2\theta}\right\} = \frac{\sqrt{1-\rho^2[1]}}{\pi(1-\rho[1]\sin 2\Theta)}. \tag{4.116}$$

Putting (4.115) and (4.116) together gives

$$p_Z(z) = \frac{\sqrt{1-\rho^2[1]}}{\pi(2z^2-2z+1)(1-\rho[1]\sin(2\arctan(1-1/z)))}. \tag{4.117}$$

This is the p.d.f. that governs the fractional zero crossing time Z, including interpolations and extrapolations. If z is only considered when a zero crossing has occurred, then interval conditioning (Peebles, 1993) can be applied to find the probability density function:

$$p_Z(z \mid \text{crossing}) = \begin{cases} \dfrac{\sqrt{1-\rho^2[1]}\left[\frac{1}{2}-\frac{1}{\pi}\sin^{-1}\rho[1]\right]^{-1}}{\pi(2z^2-2z+1)(1-\rho[1]\sin(2\arctan(1-1/z)))} & z \in [0,1] \\ 0 & \text{otherwise.} \end{cases} \tag{4.118}$$

To recapitulate: for a wide-sense stationary Gaussian process, (4.118) is the p.d.f. that describes where a linear interpolation places the zero crossing between two samples. The sole parameter in this distribution is $\rho[1]$, the correlation between successive samples.

Comparing the Analytical and Empirical Distributions

Before proceeding any further, it is prudent, as an aside, to perform a brief check to confirm that (4.117) (and hence 4.118) has been derived properly. This is accomplished by convolving white noise with a simple, seven-sample impulse response

$$h[n] = \begin{array}{ccccccc} 1 & 0.5 & -0.3 & 0.2 & -0.3 & -0.2 & -0.1, \end{array} \tag{4.119}$$

[1] using the rule $\int_0^\infty re^{-ar^2}dr = \frac{1}{2a}$.

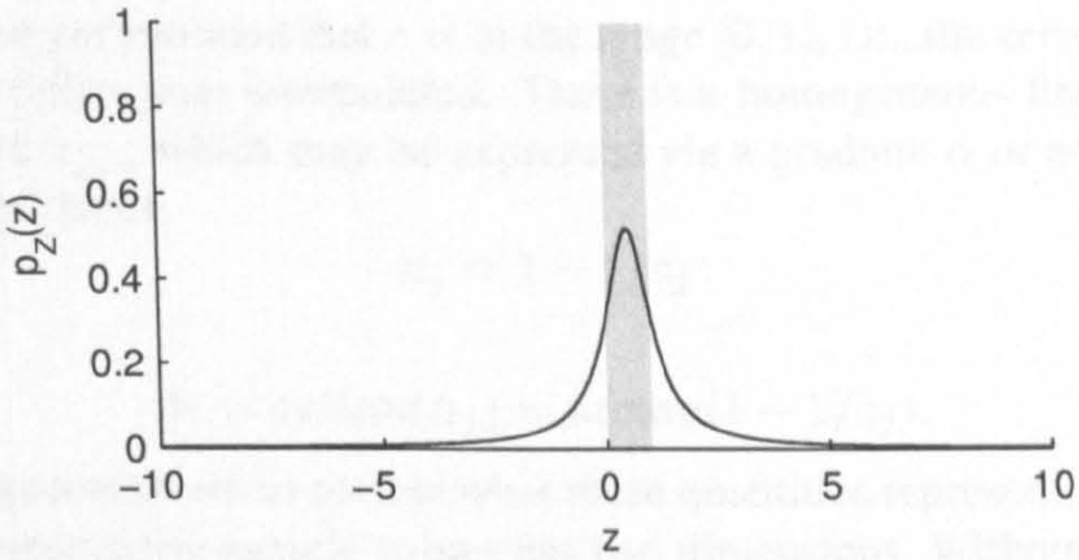


Figure 4.21: Analytical (solid) and empirical (dotted) probability density function for fractional zero crossings. The two curves are overlaid very closely. The width of the grey patch indicates the region corresponding to an interpolation.

measuring a large sample of fractional zero crossing times (permitting extrapolations), and compiling them into a histogram. Figure 4.21 overlays the histogram that results from this procedure onto the analytical probability density function computed using (4.117). The two curves are indistinguishable.

4.5.2 Probability Density Functions

Combining geometry with the bivariate Gaussian density function allowed us to arrive at the p.d.f. for fractional sample crossing time for a linear interpolation without too much difficulty. Formally, a (linear) interpolated interval we define as the random event corresponding to two successive zero crossings, I samples apart, with fractional crossing times Z_0 and Z_1. What is the probability of such an event?

First, let us designate the samples of the k-th sample of the j-th zero crossing x_{jk}. An interval consists of four samples $x_{0,0}$, $x_{0,1}$, $x_{1,0}$, $x_{1,1}$. (Refer again to the diagram in Figure 4.15.) The probability density governing these samples—maintaining the same assumptions used up to this point—is quadrivariate Gaussian:

$$p_{\mathcal{X}}(\mathbf{x} \mid H_j) = \frac{1}{4\pi^2 |\Sigma_{\mathcal{X}}|^{1/2}} \exp\left(\frac{\mathbf{x}^T \Sigma_{\mathcal{X}}^{-1} \mathbf{x}}{-2}\right) \tag{4.120}$$

where, for H_j,

$$\Sigma_{\mathcal{X}} = \begin{pmatrix} 1 & \rho_j[1] & \rho_j[i] & \rho_j[i+1] \\ \rho_j[1] & 1 & \rho_j[i-1] & \rho_j[i] \\ \rho_j[i] & \rho_j[i-1] & 1 & \rho_j[1] \\ \rho_j[i+1] & \rho_j[i] & \rho_j[1] & 1 \end{pmatrix}. \tag{4.121}$$

The fractional crossing time for the j-th zero crossing is

$$z_j = \frac{x_{j,0}}{x_{j,0} - x_{j,1}}. \tag{4.122}$$

Note that it is *not* yet assumed that z is in the range $[0, 1]$, i.e., the zero crossing might be extrapolated rather than interpolated. There is a homogeneous linear relationship between $x_{j,0}$ and $x_{j,1}$, which may be expressed via a gradient α or angle θ, as in the previous section. Hence

$$\alpha_j = 1 - 1/z_j \tag{4.123}$$

and

$$\theta_j = \arctan(\alpha_j) = \arctan(1 - 1/z_j). \tag{4.124}$$

It is worth taking a moment to review what these quantities represent. The probability space for two consecutive sample values has two dimensions. Without any knowledge concerning what these values are, the event can lie anywhere in this space. If it is at least known that a zero crossing has occurred, then the samples have opposite sign and the event must lie in quadrant II or IV. (For the sampled interval detector, this is the extent of the crossing information conveyed by the test statistic.)

If the interpolated zero crossing time is available, then the first sample is proportional to the second by a constant α, determined from the crossing time. This means that the random event must lie along a line with gradient α. Because the function relating z to α is smooth for all values $z \neq 0$, it follows that a small interval in z corresponds to a small interval in α; furthermore, an interval in α, is an interval in θ because arctan is a smooth function. Therefore, intuitively, a small region in z is a small (wedge-shaped) region in probability space, and the problem of finding the probability of z reduces to the problem of transforming differential areas in the Gaussian density function.

In the case of two interpolated zero crossings, there is a four-dimensional space: two dimensions belonging to each crossing. If the both zero crossing times are known, then the random event must lie on two lines, one passing through the first pair of dimensions, the other passing through the second pair of dimensions. In terms of differential areas, the probability of a small region $[z_0, z+\delta z_0]$, $[z_1, z+\delta z_1]$ transforms to wedge-shaped regions in four dimensions $[\theta_0, \theta+\delta\theta_0]$, $[\theta_1, \theta+\delta\theta_1]$.

The first step is to rotate the space $\mathcal{X}$ in such a way that the events align with the axes. The scale-invariant rotation which achieves this is associated to the block diagonal matrix

$$T_1 = \begin{pmatrix} \cos\theta_0 & \sin\theta_0 & 0 & 0 \\ -\sin\theta_0 & \cos\theta_0 & 0 & 0 \\ 0 & 0 & \cos\theta_1 & \sin\theta_1 \\ 0 & 0 & -\sin\theta_1 & \cos\theta_1 \end{pmatrix}. \tag{4.125}$$

Transforming the probability space $\mathcal{X}$ using T_1 gives a new probability space $\mathcal{Y}$, where

$$\mathbf{y} = T_1\mathbf{x} \tag{4.126}$$

and the elements of the random vector $\mathbf{y}$ are joint Gaussian-distributed, with zero mean and covariance matrix (Peebles, 1993; Whalen, 1971)

$$\Sigma_{\mathcal{Y}} = T_1\Sigma_{\mathcal{X}}T_1^T. \tag{4.127}$$

Random events corresponding to a pair of interpolated crossings at $\langle z_0, z_1 \rangle$ lie on the plane $y_{0,1} = y_{1,1} = 0$. To find the probability mass associated with a small wedge

surrounding this plane, it is necessary to transform the space $\mathcal{Y}$ into a space $\mathcal{R}$ defined by a pair of polar co-ordinate systems. The transformation is accomplished by

$$y_{0,0} = R_0 \cos\phi_0 \tag{4.128}$$
$$y_{0,1} = R_0 \sin\phi_0 \tag{4.129}$$
$$y_{1,0} = R_1 \cos\phi_1 \tag{4.130}$$
$$y_{1,1} = R_1 \sin\phi_1 \tag{4.131}$$

and has Jacobian $J = |R_0 R_1|$. The probability mass associated with the observed interpolated zero crossing times lies along the plane with polar angles $\phi_0 = 0, \pi$ and $\phi_1 = 0, \pi$ (alternatively expressed in $\mathcal{Y}$-space above). Point conditioning on $\phi_0{=}\phi_1{=}0$ yields a probability density function in R_0 and R_1, which measure the distance along the wedges, for the case where the first sample in each crossing is *positive*. (We shall return to the alternative cases shortly.) This gives a p.d.f. in $\mathcal{R}$-space

$$p_{\mathcal{R}}(R_0, R_1, \phi_0 = 0, \phi_1 = 0) = \frac{1}{4\pi^2|\Sigma_{\mathcal{X}}|^{1/2}} R_0 R_1 \exp\left(\frac{\kappa_{1,1}R_0^2 + \kappa_{3,3}R_1^2 + 2\kappa_{1,3}R_0R_1}{-2}\right). \tag{4.132}$$

where $\kappa_{i,j}$ is the (i,j)-th element of $\Sigma_{\mathcal{Y}}^{-1}$. At this point it becomes clear why it was desirable to rotate the $\mathcal{X}$-space at an earlier stage to align the events with the axes. All the terms in the exponent following conversion to polar co-ordinates contain products of two trigonometric functions, e.g., $\cos\phi_0 \sin\phi_1$, $\sin\phi_0 \cos\phi_0$. Then, because the conditioning is on $\phi_j = 0$, every term containing a sine function disappears, leaving only those in (4.132) whose coefficients are κ's.

The next step is to turn (4.132) into a quantity independent of R_0 and R_1. Considering once again the signal from which these quantities were derived, we see that the R_j relate to the samples by $R_j^2 = x_{j,0}^2 + x_{j,1}^2$; in other words, the R_j convey the root mean square of the zero crossing samples. In a timing-only scheme, these quantities are unknown and must be marginalised by performing the integration

$$p_{\mathcal{R}}(\phi_0 = 0, \phi_1 = 0) = \frac{1}{4\pi^2|\Sigma_{\mathcal{X}}|^{1/2}} \int_0^\infty \int_0^\infty R_0 R_1 \exp\left(\frac{\kappa_{1,1}R_0^2 + \kappa_{3,3}R_1^2 + 2\kappa_{1,3}R_0R_1}{-2}\right) dR_0 dR_1. \tag{4.133}$$

This integration is made difficult by the coupling of R_0 and R_1 in the exponential (i.e., the existence of the $2\kappa_{1,3}R_0R_1$ term). Employing an eigenvalue decomposition to decouple terms in a Gaussian exponential is standard practice (see, e.g., Peebles (1993); Shanmugan and Breipohl (1988); Whalen (1971)) and is demonstrated next. Employing the abbreviations

$$Q = \begin{pmatrix} \kappa_{1,1} & \kappa_{1,3} \\ \kappa_{1,3} & \kappa_{3,3} \end{pmatrix} \text{ and } \mathbf{r} = \begin{pmatrix} R_0 \\ R_1 \end{pmatrix} \tag{4.134}$$

we can write (4.134) as

$$p_{\mathcal{R}}(\phi_0 = 0, \phi_1 = 0) = \frac{1}{4\pi^2|\Sigma_{\mathcal{X}}|^{1/2}} \int_0^\infty \int_0^\infty R_0 R_1 \exp\left(\frac{\mathbf{r}^T Q \mathbf{r}}{-2}\right) dR_0 dR_1. \tag{4.135}$$

To decouple (and rescale) the variables it is necessary to find a transform T_2 so that

$$(T_2\mathbf{r})^T Q(T_2\mathbf{r}) = \mathbf{r}^T(T_2^T Q T_2)\mathbf{r}$$

and

$$T_2^T Q T_2 = I \tag{4.136}$$

where I is the 2×2 identity matrix. Let U and D respectively denote a column matrix of eigenvectors and a diagonal matrix of eigenvalues for Q so that $QU = DU$ and

$$U^{-1}QU = D. \tag{4.137}$$

Using $D = \sqrt{D}\sqrt{D}$, we can write

$$\sqrt{D^{-1}}U^{-1}QU\sqrt{D^{-1}} = I, \tag{4.138}$$

then because i) U is an orthogonal matrix, $U^{-1} = U^T$ and ii) D is a diagonal matrix, $D^{-1} = (D^{-1})^T$, it follows that

$$\left(\sqrt{D^{-1}}\right)^T U^T QU\sqrt{D^{-1}} = \left(U\sqrt{D^{-1}}\right)^T QU\sqrt{D^{-1}} = I. \tag{4.139}$$

Using this result in (4.136), it is seen that the transform required to decouple the exponentiated variables is $T_2 = U\sqrt{D^{-1}}$. In the working that follows, it will prove useful to label the individual elements of T_2 using

$$T_2 = \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \tag{4.140}$$

The space resulting from this transform is designated $\mathcal{R}'$. In this space, the integration has the formulation

$$p_{\mathcal{R}'}(\phi_0 = 0, \phi_1 = 0) = \tag{4.141}$$
$$\frac{ad - bc}{4\pi^2|\Sigma_{\mathcal{X}}|^{1/2}} \iint (aR'_0 + bR'_1)(cR'_0 + dR'_1) \exp\left(\frac{R'^2_0 + R'^2_1}{-2}\right) dR'_0 dR'_1.$$

This integral is readily transformed into polar co-ordinates using the transformed variables

$$R'_0 = \Gamma\cos\Phi \tag{4.142}$$
$$R'_1 = \Gamma\sin\Phi. \tag{4.143}$$

with Jacobian $|\Gamma|$, whereupon the angular bounds become

$$\Omega_0 = \angle\left((1,i)T_2^{-1}(0,1)^T\right) \tag{4.144}$$
$$\Omega_1 = \angle\left((1,i)T_2^{-1}(1,0)^T\right). \tag{4.145}$$

The final integration is performed in Γ-space, in which the variables may be separated completely

$$p_{\Gamma}(\phi_0 = 0, \phi_1 = 0) = \tag{4.146}$$
$$\frac{ad - bc}{4\pi^2|\Sigma_{\mathcal{X}}|^{1/2}} \int_{\Omega_0}^{\Omega_1} (a\cos\Phi + b\sin\Phi)(c\cos\Phi + d\sin\Phi)d\Phi \int_0^{\infty} \Gamma^3 e^{\frac{\Gamma^2}{-2}} d\Gamma.$$

which has the closed-form solution

$$\frac{ad-bc}{4\pi^2|\Sigma_{\mathcal{X}}|^{1/2}}\left[\frac{(ac-bd)\sin 2\Phi}{2}+(ac+bd)\Phi+(ad+bc)\sin^2\Phi\right]_{\Omega_0}^{\Omega_1}. \quad (4.147)$$

Here, (4.147) is the probability density function for θ_0 and θ_1 in the region where both are positive. This is equivalent to the probability mass associated with two crossings whose first sample is positive.

As it stands, (4.147) provides only limited information. Ultimately, we want to know the probability of two fractional zero crossings z_0 and z_1, spaced i samples apart, regardless of their direction, that is, $p(z_0, z_1 \mid i)$. The result is found to be

$$p(z_0, z_1 \mid i) = \frac{2p_{\mathcal{R}}(\phi_0=0,\phi_1=0)+2p_{\mathcal{R}}(\phi_0=0,\phi_1=\pi)}{\left(\frac{1}{2}-\frac{1}{\pi}\sin^{-1}\rho[1]\right)(2z_0^2-2z_0+1)(2z_1^2-2z_1+1)}. \quad (4.148)$$

The term $p_{\mathcal{R}}(\phi_0{=}0, \phi_1{=}0)$ is the probability of two zero crossings whose first samples are positive; the probability is the same if the first samples are *negative*, so it is doubled. The term $p_{\mathcal{R}}(\phi_0{=}0, \phi_1{=}\pi)$ is the probability that the first sample of the first crossing is positive and the first sample of the second crossing is negative. This is computed by proceeding from (4.134) and negating $\kappa_{1,3}$. This quantity is also doubled to account for the reverse scenario. Finally, the factors in the denominator are familiar: the first factor conditions on a crossing event; the second two factors result from the Jacobian determinant for replacing $\langle z_0, z_1\rangle$ with $\langle\theta_0, \theta_1\rangle$.

Comparing the Analytical and Empirical Distributions

Formulae are now available for the joint probability density of two zero crossing angles θ_j and two zero crossing fractional times z_0 and z_1, given the separation of i samples between the crossings. It is worth checking that these probability density functions resemble histograms generated using random data. The angles and fractional times are computed for white Gaussian noise convolved with the seven-sample impulse response[1]

$$h[n] = \begin{array}{ccccccc} 1 & 0.5 & -0.3 & 0.2 & -0.3 & -0.2 & -0.1, \end{array} \quad (4.149)$$

when the samples are separated by $i = 6$. The two-dimensional probability densities are shown in Figure 4.22 as images. The density function for two crossings angles is shown in two formats: the first (A) uses the four-quadrant inverse tangent function, implying some knowledge of the sign of the samples, i.e., $\theta_j = \text{atan2}(x_{j,1}, x_{j,0})$; the second (B) uses the principal value of the angle computed from the zero crossing time, $\arctan(1-1/z_j)$. Figures 4.22C and 4.22D respectively show the analytical and empirical probability density for two fractional crossing times. The visual agreement between the two images suggests that the analytical approach works. All that remains is to incorporate this procedure into the decision rule of the interpolated interval detector and to repeat the evaluation experiment a final time.

[1] A narrowband filter impulse response produces a two-dimensional probability density function that appears as a thin sliver when plotted as a two-dimensional image, and this makes it difficult to assess visually whether the empirical and analytical p.d.f.s match. For this reason, an impulse response that weakly correlated the samples was used. It is the same impulse response as (4.119).

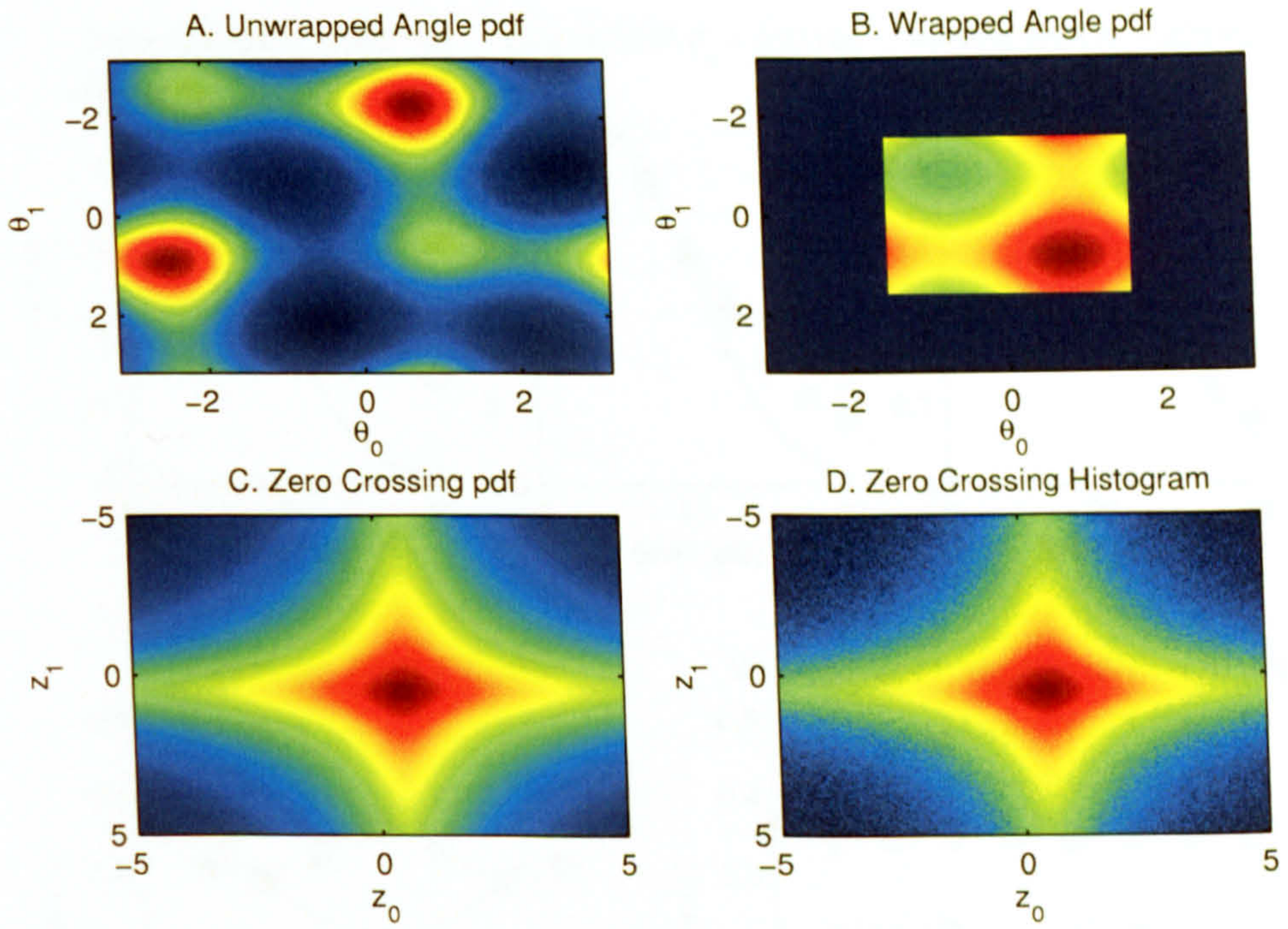


Figure 4.22: Probability density function for a pair of interpolated crossings (see text). A) the p.d.f. for two angles; B) the p.d.f. for two principal-valued angles; C) the p.d.f. for two zero crossing fractional times; D) the histogram corresponding to (C).

4.5.3 Setting up the Experiments

The narrowband detection experiment described in Section 4.1 is repeated once again for the interpolated interval detector. One of the main interests, as with the continuous interval detector, is to see whether the interpolated interval detector overcomes the problem of performance degradation at high frequencies by incorporating additional information from the fractional zero crossings.

The procedure for finding the likelihood functions for H_0 and H_1 is straight-forward. The autocorrelation function for each system is found in the same manner as the sampled interval detector. As each interval is received, the values for i, z_0 and z_1 are computed and submitted to the function (4.147) conditioned on $\rho_0[k]$ and $\rho_1[k]$. Whichever hypothesis associates the most probability density with the measurements is selected. An alternative and equivalent implementation of this likelihood test operates on i, θ_0 and θ_1.

In the three earlier experiments, little effort was required to predict the probability of error before experimental results were obtained. Predicting the performance of the interpolated interval detector is substantially more time-consuming because the likelihood functions must be integrated effectively in three dimensions rather than one. For this reason, only the experimental probability of error is considered.

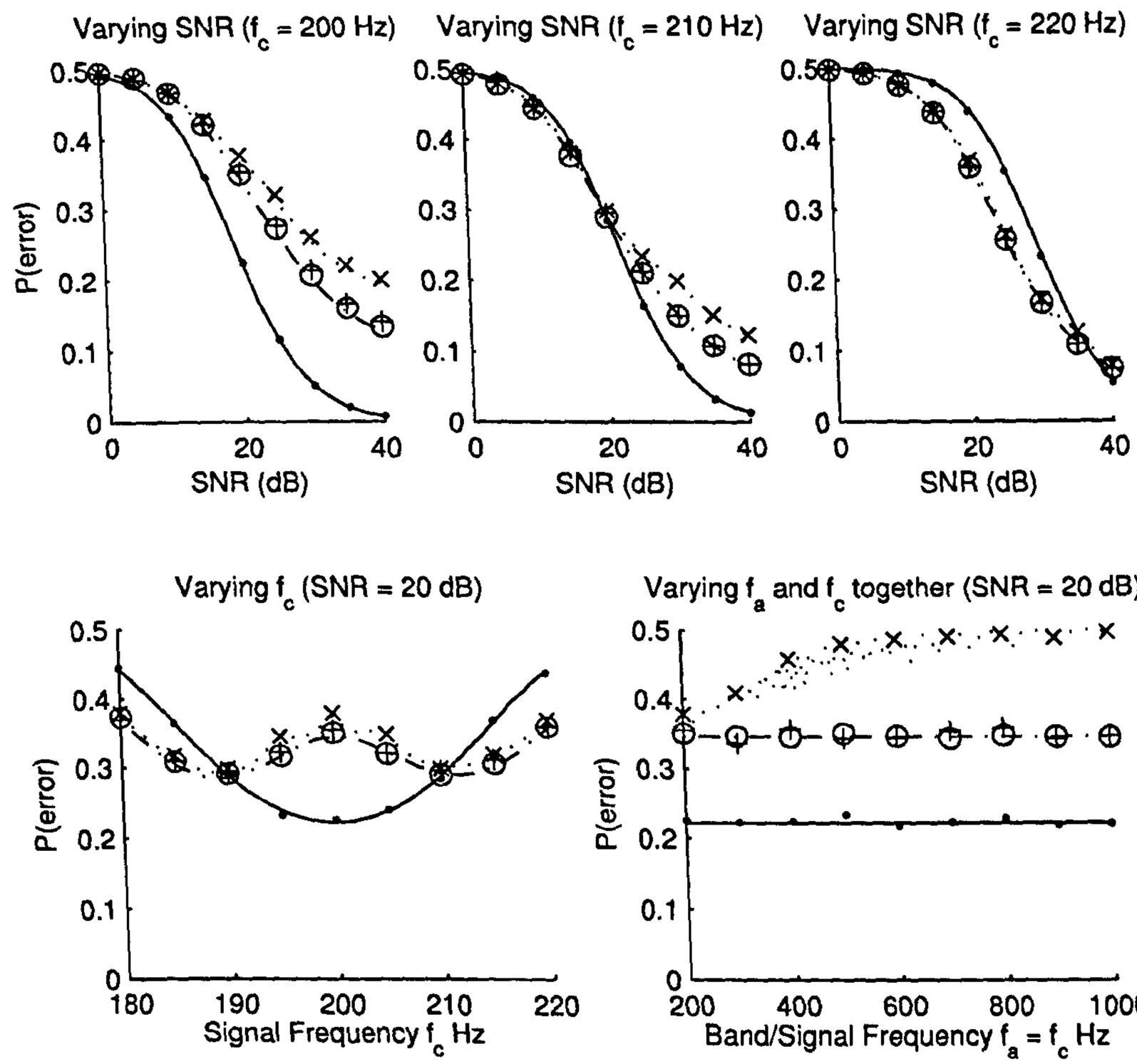


Figure 4.23: Probability of error in the detectors presented so far. The predicted and observed values are shown using lines and markers, according to the following key: squared-envelope detector (solid line; solid circle •); sampled interval detector (dotted; cross ×); continuous interval detector (dash-dotted; open circle ○) and interpolated interval detector (no analytical results; plus +).

4.5.4 Experimental Results and Analysis

Figure 4.23 reveals a close correspondance between the results of the continuous interval detector and those of the interpolated interval detector. We shall not, therefore, explicitly address the first three research questions advanced in Section 4.1.5, except to note that the answers are identical to those given for the continuous interval detector in Section 4.18.

Is it possible to predict the performance of the detector? It is presumably possible to predict the performance of the interpolated interval detector by a fine-grained numerical integration of the probability density functions. The main intention behind this question was to ensure that the theory and experimental results were in accordance, as a means

of mutual validation. The close match between the results of the CID and IID serves to validate the latter.

How do the results of the different detectors compare? The similarity between the performance of the continuous and interpolated interval detector is expected, given that both detectors have access to the similar information, expressed in a slightly different form. Concerning the small differences in performance where they do appear: at most data points, the interpolated interval detector performs slightly *worse* than the continuous interval detector (e.g., the top left-hand plot of Figure 4.23). It is rather surprising that the performance of the IID is worse than that of the CID, considering that i) the IID ideally commits no model errors, and the CID certainly does; and ii) the continuous interval statistic, i_c, provides less information than the test statistic $\langle i, z_0, z_1 \rangle$.

There are at least two reasons why the probability of error estimated for the interpolated interval detector might appear higher than that given for the continuous interval detector. First, these results were obtained by measuring the outcome of a finite number of trials—a standard approach to estimating population statistics. If too few trials were conducted (i.e., the detectors were evaluated in too few instances) then it is possible that the results did not converge adequately, and that, given further trials, the IID might be shown eventually to outperform the CID. A second possibility—and the more likely, in this author's opinion—is that numerical instability in the linear transformations required to obtain the IID likelihood functions (*cf.* Section 4.5.2) led to a mild degradation in performance. (Specifically, the matrix $\Sigma_{\mathcal{Y}}$ is badly-conditioned for narrowband processes.)

4.6 Summary

This chapter has developed three elementary interval detectors and evaluated each one in a simple task requiring the detection of a narrowband Gaussian process against a background of white noise. Each detector extracts one zero crossing interval from the output of an analysis filter and then decides on the basis of this measurement whether or not the signal is present.

The sampled interval detector chooses between a signal-and-noise and noise-only model on the basis of the number of samples between two zero crossings. This detector was generally only effective at low frequencies where the intervals were sufficiently sampled to resolve frequency information. For a sample rate of 16384 Hz, the detector performed poorly above about 200 Hz.

The continuous interval detector was developed as the first solution to the sampling problem and uses continuous signal models in conjunction with interpolated zero crossings. The detector has been shown to perform consistently at all frequencies tested (200 Hz–1 kHz).

Another solution to the sampling problem was the interpolated interval detector, which uses a coarse measure of the interval, like the sampled interval detector, but augments the detection decision with additional information from the fractional zero crossing times computed by a linear interpolation. This detector achieved a similar probability of error to the continuous interval detector.

At each stage, the performance of the interval detectors was compared with that of a squared-envelope detector. The two were shown to behave quite differently. The envelope detection is dependent exclusively upon the attenuation of the signal and the SNR, whereas interval detection is sensitive to the frequencies and bandwidths of the signals concerned. If the centre of the analysis filter and the signal frequency coincide, then detections must be made solely on the basis of interval variance; if the signal is displaced from the band centre, the mean interval is affected and detection improves.

Chapter

5

Further Developments of the Interval Detector

In order to place this chapter in its context, it is helpful to retrace the steps that have led up to this point. In Chapter 2, we reviewed some of the leading theories concerning the encoding of acoustic signals in the auditory nerve and provided a survey of computational models that generate auditory-style signal representations. Chapter 3 opened with a short account of conventional, power-based sonar detection and proceeded to investigate whether spectrograms generated by auditory-motivated algorithms could be adapted to narrowband sonar applications. It was difficult to judge the superiority of one algorithm over another simply by viewing the spectrograms, and it seemed evident that an evaluation based on statistical detection theory would deliver more robust conclusions. To this end Chapter 4 considered the most elementary unit of information in most timing-based representations—the zero crossing interval—and developed an optimal detector to operate on this test statistic.

Looking ahead, our ultimate intention is to develop auditory-motivated algorithms to detect, track and group tonal components in real sonar signals, and ideally this development should be guided by a comprehensive account of the statistics of temporal representations, such as the EIH and ZCPA. However, there remains a gulf between the modest results of Chapter 4 and the full-blown description implied above. The purpose of Chapter 5 is to bridge this gap by extending the detection routines of Chapter 4 in a number of useful directions. These extensions each attempt to remove a restriction imposed at the beginning of the previous chapter.

First, the probability of error was the sole performance metric, whereas conventional sonar system performance analysis more often considers the probability of detection when the false alarm rate is held fixed (Burdic, 1984). The decision rule of the interval detector must be modified to maximise the probability of detection rather than to minimise the probability of error. This will allow us to present the performance of an interval detector in a more familar way (e.g., via ROC curves).

Second, we assumed that the signal was a stationary Gaussian random process. The most appropriate model for a clean tonal is a sinusoid with constant amplitude and frequency, and random phase. This kind of process is *non-Gaussian* and therefore demands separate treatment.

Third, we assumed that the detectors operated on timing information alone; all information concerning the envelope was discarded. By contrast, the auditory system appears to preserve information about timing *and* power, as do models such as the EIH and ZCPA. The results from Chapter 4 demonstrate that the power detector is superior if the signal in placed near the band centre, but the interval detector is superior if the signal is sufficiently displaced from the band centre. This clearly motivates the search for a hybrid detector, which capitalises on the information in both the zero crossing intervals and the envelope.

Fourth and fifth, we assumed that the detector operated on a single interval recorded from a single analysis filter. Power-based sonars integrate information over long time periods, in a process referred to as *post-detection integration*, in order to secure a high performance at a low signal-to-noise ratio. If timing-only or hybrid detectors are to compete at similar SNRs, we will need to understand how to combine information from many zero crossings as they arrive. The topic of detection using multiple intervals on short times scales falls under a single heading (§5.4). Detection using intervals that are sufficiently removed in time to be independent and identically-distributed is discussed under a separate heading (§5.5).

Chapter 5 Outline

Section
Performance Metrics (5.1)
Detection of a Sinusoid (5.2)
Combining Power and Timing Detection (5.3)
Detection using Multiple Intervals (5.4)
Post-detection Integration (5.5)
Summary (5.6)

5.1 Performance Metrics

The probability of error (PE) figured prominently in Chapter 4, in at least three senses. First, the PE was the *target quantity* that the decision rule of every detector was designed to minimise. Second, the probability of error was the only *dependent variable* considered. In each experiment, a parameter such as the SNR or signal frequency was varied, and the effect upon a detector's PE was recorded in the form of a graph (e.g., Figures 4.5, 4.12, 4.18 and 4.23). And third, the PE was the sole *evaluation metric* for comparing the performance of detectors.

Detectors that minimise the probability of error assign an equal cost to false alarms and false dismissals. In many sonar applications, however, it is desirable to assign a different cost to each kind of error, usually in order to maintain a constant false alarm rate (CFAR). The false alarm probability depends entirely on the noise-only likelihood function. Finding the optimal decision regions for a power detector is straight-forward, because in most circumstances, the addition of a signal consistently raises the average received power. At its simplest, a power detector need only threshold any test statistic that varies monotonically with power, such as the envelope or squared-envelope.

Incorporating CFAR into interval detection is less obvious. Adding a narrowband signal to band-pass noise causes the zero crossing intervals to gravitate towards a central value related to the signal frequency. This in turn implies (at least) *two* thresholds: one threshold rejects intervals that are too short, the other rejects intervals that are too long. It is possible to set the probability of false alarm by sliding the first threshold to adjust the sensitivity of the detector to short intervals. Alternatively, one could modify the second threshold and adjust the sensitivity of the detector to long intervals. Either approach will affect the probability of detection differently, and the question naturally arises: What combination of thresholds fixes the probability of false alarm at the desired value *and* maximises the probability of detection? To answer this question, we must revisit the Neyman-Pearson criterion.

It was seen in Chapter 3 that the Neyman-Pearson criterion fixes the probability of false alarm at a prespecified value, p_{fa}, and then proceeds to maximise the probability of detection. (The prior probabilities are assumed to be uniform.) The Neyman-Pearson criterion is satisfied by the decision rule

$$\text{choose } H_1 \text{ iff } \frac{p_{I_c}(i_c \mid H_1)}{p_{I_c}(i_c \mid H_0)} > \eta, \text{ otherwise choose } H_0. \tag{5.1}$$

The decision regions can be calculated for a variety of η by inserting the appropriate conditional interval probability density functions into (5.1). The values of i_c that cause the likelihood ratio to exceed the likelihood threshold are assigned to $\mathcal{R}_1$; the remainder are assigned to $\mathcal{R}_0$. Once the decision regions have been determined, the probability of detection and false alarm may be found by integration.

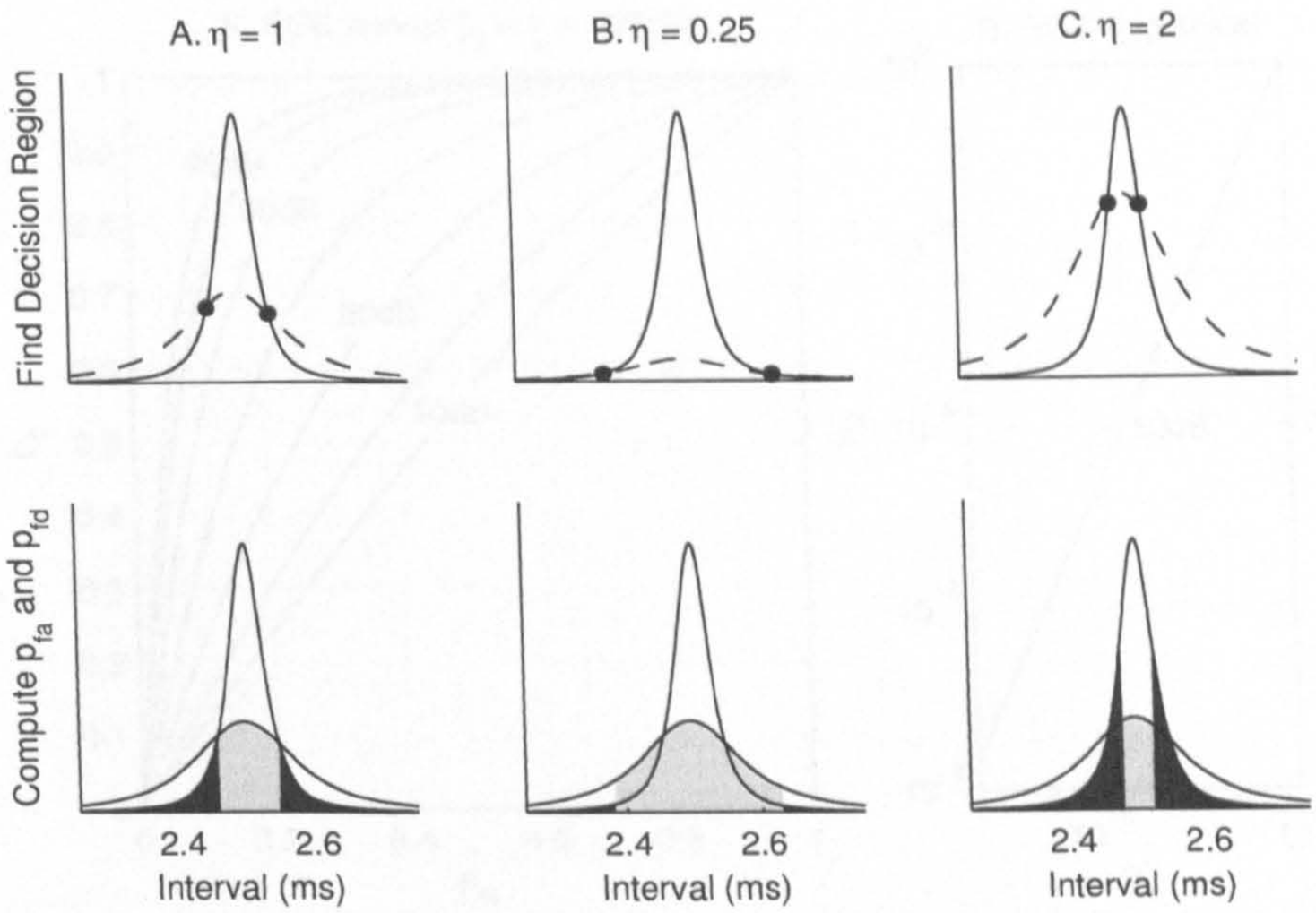


Figure 5.1: Decision regions for an interval detector computed using MGMMs. The area of the light grey region corresponds to p_{fa}; the dark grey area corresponds to the probability of false dismissal (p_{fd}), or equivalently, $1 - p_d$. The parameter η adjusts the trade-off between false alarms and false dismissals. A) minimum error decision regions; B) lowered likelihood threshold, fewer false alarms, increased specificity; C) raised likelihood threshold, fewer false dismissals, increased sensitivity.

5.1.1 Examining the Decision Regions

The decision regions of interval and power detectors are qualitatively different. Most importantly, the likelihood functions of a squared-envelope detector intersect at exactly one point—provided that there is some difference in power between the two hypotheses—whereas those of a continuous interval detector generally intersect at two points. This restates the earlier observation that adding a narrowband signal to noise (primarily) *increases the mean* of the squared-envelope but *decreases the spread* of the zero crossing intervals.

Figure 5.1A marks the intersections of a noise-only and signal-and-noise interval p.d.f. when η is set to one. One can see from (5.1) that varying η effectively rescales the noise-only likelihood function: decreasing η causes the intersections to move apart (Figure 5.1B), and increasing η causes the intersections to close together (Figure 5.1C). The corresponding effect upon the probability of false dismissal (i.e., $1 - p_d$) and probability of false alarm (p_{fa}) is conveyed by the lower row of plots of Figure 5.1. Decreasing η causes the probability of false alarm (light shaded region) to increase and the probability of false dismissal (dark shaded region) to decrease, and *vice versa*.

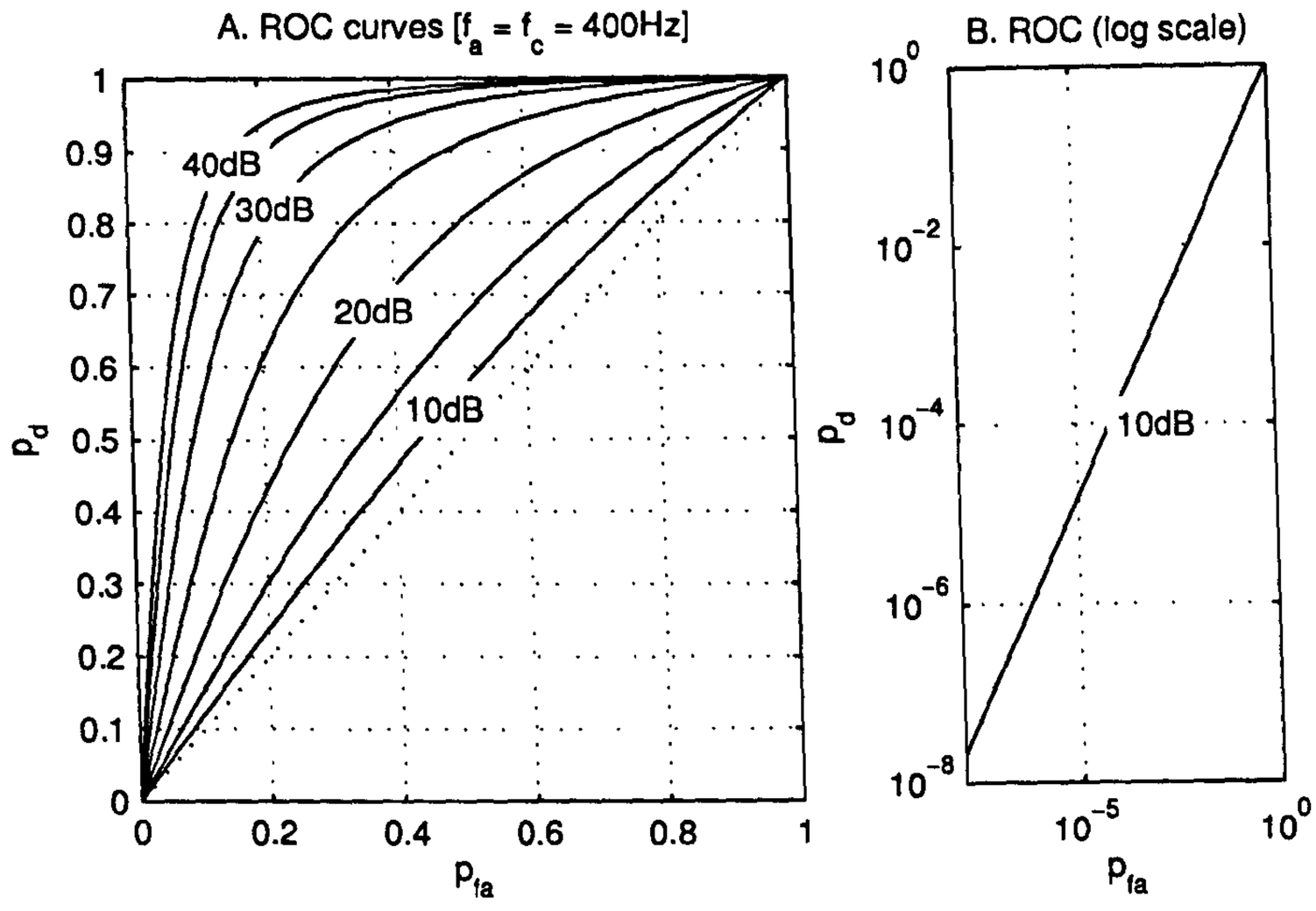


Figure 5.2: A) receiver operating characteristic (ROC) curves for the continuous interval detector for signal-to-noise ratios from 10 dB to 40 dB in 5 dB steps; B) ROC curve relating detection performance at 10 dB SNR, plotted on a log-log scale.

5.1.2 Producing ROC Curves for Interval Detectors

To produce a ROC curve for the continuous interval detector, we repeat the following two steps for a variety of η: i) compute the decision regions $\mathcal{R}_0$ and $\mathcal{R}_1$ by finding the intersections of $\eta p_{I_c}(i_c \mid H_0)$ and $p_{I_c}(i_c \mid H_1)$; ii) determine the true positive and false positive probabilities by integration.

$$p_d = \int_{\mathcal{R}_1(\eta)} p_{I_c}(i_c \mid H_1) di_c \tag{5.2}$$

$$p_{fa} = \int_{\mathcal{R}_1(\eta)} p_{I_c}(i_c \mid H_0) di_c. \tag{5.3}$$

There are at least three implementational issues to consider here: first, how to choose the set of η's; second, how to find the intersections between the p.d.f.s; and third, how to perform the integrations. The MATLAB script written to generate the figures in this section chooses the η's in two steps: the first, "coarse" step distributes 100 points evenly on a log scale between 10^{-2} and 10^{2}, and a second, "refinement" step then interpolates 100 points into the regions of the ROC curve that are least smooth, i.e., where the differences in p_{fa} or p_d are greatest. The script performs the second and third stages numerically: the conditional p.d.f.s are discretised, the decision regions are determined by point-wise comparisons, incorporating the η factor, and the integrations are replaced by summations with an integration step of 0.1 μs. (This procedure directly corresponds to that suggested by Figure 5.1.)

Figure 5.2A shows a set of ROC curves for a continuous interval detector. The signal is a random process formed by convolving white noise with the impulse response with MGMM definition

$$\Lambda_{h_s} = \langle A_1, C = 2.5, \mu = 0, \bar{\omega} = 2\pi \cdot 400, \phi = 0 \rangle \tag{5.4}$$

$$+ \langle A_1, C = 2.5, \mu = 0, \bar{\omega} = -2\pi \cdot 400, \phi = 0 \rangle , \tag{5.5}$$

where A_1 is included to scale the process to unit power. The signal must be detected against a background of white noise with power spectral density $N_0/2$. The impulse response of the analysis filter has the MGMM definition

$$\Lambda_{h_a} = \langle A = 1, C = 40, \mu = 0, \bar{\omega} = 2\pi \cdot 400, \phi = 0 \rangle \tag{5.6}$$

$$+ \langle A = 1, C = 40, \mu = 0, \bar{\omega} = -2\pi \cdot 400, \phi = 0 \rangle . \tag{5.7}$$

To produce an ROC curve in Figure 5.2, a signal-to-noise ratio is chosen, and the true positive and false positive probabilities are computed analytically for two-hundred values of η, spaced logarithmically between 10^{-2} and 10^2 (see above). A set of ROC curves is obtained by varying the signal-to-noise ratio.

A ROC curve provides an overall picture of a detector's performance at a particular SNR. The top-left hand corner $(0, 1)$ is the perfect classifier; the points on the dotted line, $p_d = p_{fa}$, are equivalent to chance performance. For example, the probability of detecting a signal when the signal-to-noise ratio is 30 dB, constraining the probability of false alarm to 0.1, is read off from the graph as 0.52. Similarly, to secure a false alarm probability of 0.2 and a detection probability of 0.4, an SNR of approximately 20 dB is required. For very small p_{fa}, it is often appropriate to plot the ROC curve on a log-log or log-linear scale. Figure 5.2B shows the ROC for the 10 dB SNR condition plotted on a log-log scale. The interval probability density functions are imperfect in the tail due to ill-conditioning (*cf.* Section 4.3.3); consequently, obtaining an estimate of p_d for very low p_{fa} is often impossible unless the SNR is very low. For this reason, the ROC curves included here are plotted on a linear scale.

Computing ROC Curves for the Envelope Detector

The continuous interval detector ROC curves were determined implicitly by computing (p_{fa}, p_d) pairs, whilst varying the free parameter η. We turn now to the production of ROC curves for the baseline power detector, which operates on a single sample of the squared-envelope. In this case, an explicit function $p_d(p_{fa})$ is readily available. It can be shown that two exponential probability density functions with $\sigma_1^2 > \sigma_0^2$ intersect at exactly one point, $e = \epsilon$, and that $\mathcal{R}_1 = \{e : e > \epsilon\}$. The probability of false alarm is therefore given by

$$p_{fa} \equiv P(D_1 \mid H_0) = \int_{\epsilon}^{\infty} \frac{1}{2\sigma_0^2} \exp\left(\frac{e}{-2\sigma_0^2}\right) de, \tag{5.8}$$

which does not depend on the signal-and-noise distribution. The threshold on the test statistic, ϵ, can be computed for a given probability of false alarm by solving and rearranging (5.8), i.e.,

$$\epsilon = -2\sigma_0^2 \log p_{fa}. \tag{5.9}$$

Finally, placing (5.9) into the expression for the probability of detection, we arrive at an expression for the relationship between p_{fa} and p_d, which can be used to construct a ROC curve directly.

$$p_d \equiv P(D_1 \mid H_1) = \int_{\epsilon}^{\infty} \frac{1}{2\sigma_1^2} \exp\left(\frac{e}{-2\sigma_1^2}\right) de \tag{5.10}$$

$$= \exp\left(\frac{\sigma_0^2}{\sigma_1^2} \log p_{fa}\right) \tag{5.11}$$

$$= p_{fa}^{\sigma_0^2/\sigma_1^2}. \tag{5.12}$$

Figure 5.3 plots four sets of ROC curves for the squared-envelope and continuous interval detectors, in which the signal frequency is respectively set to 400, 410, 420 and 430 Hz. The narrowband SNR is the family parameter and assumes the values 10, 20, 30 and 40 dB.

5.1.3 Identifying Regions of Superior Performance

It is difficult to gain an overall insight into the conditions under which the continuous interval detector outperforms the squared-envelope detector, simply by examining the collection of ROC curves in Figure 5.3. The family parameter in a set of ROC curves must be discretised in order to present the data as a series of curves—or on separate axes altogether—whereas one is often interested to know how performance is affected as a parameter is varied *smoothly*. With this in mind, we shall define a superiority measure, p'_d, which relates the benefit of using an interval detector, if any there is any, in terms of the increase in detection probability:

$$p'_d(\mathbf{c}) = \begin{cases} p_{d,\text{cid}}(\mathbf{c}) - p_{d,\text{sed}}(\mathbf{c}) & p_{d,\text{cid}}(\mathbf{c}) > p_{d,\text{sed}}(\mathbf{c}) \\ 0 & \text{otherwise.} \end{cases} \tag{5.13}$$

Here, $\mathbf{c}$ abbreviates a set of conditions, e.g., p_{fa}, SNR and f_c; and $p_{d,\text{cid}}(\cdot)$ and $p_{d,\text{sed}}(\cdot)$ refer to the corresponding probability of detection for the continuous interval and squared-envelope detector, respectively. Note that p'_d is zero whenever the interval detector is inferior to the squared-envelope detector.

Each image in the top row of Figure 5.4 displays a two-dimensional representation of $p'_d(p_{fa}, \text{SNR})$, with f_c held fixed at five different frequencies. Grey areas highlight the regions in which an interval detector is superior, with brighter regions indicating larger increases in performance[1]. The black areas represent regions of inferior performance. Similarly, each image in the bottom row of Figure 5.4 displays a two-dimensional

[1] It should be noted that the values in each image of Figure 5.4 are scaled to occupy the full colour map in order to maximise contrast. The peak values of p'_d across each image are tabulated below. The layout of the cells corresponds to the layout of the subfigures.

0.0076	0.0555	0.1239	0.1891	0.2450
0.0059	0.0453	0.1538	0.2442	0.1009

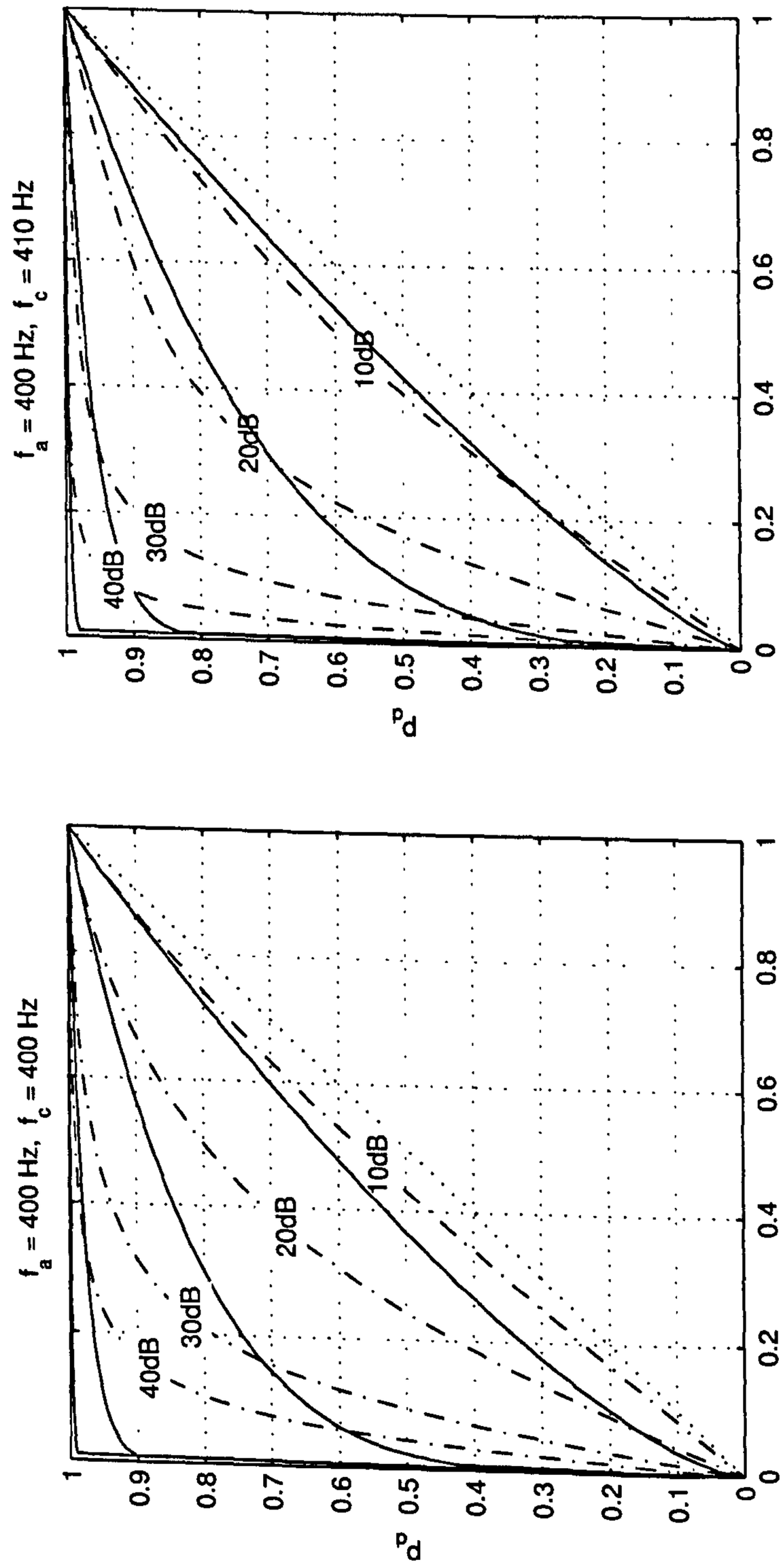
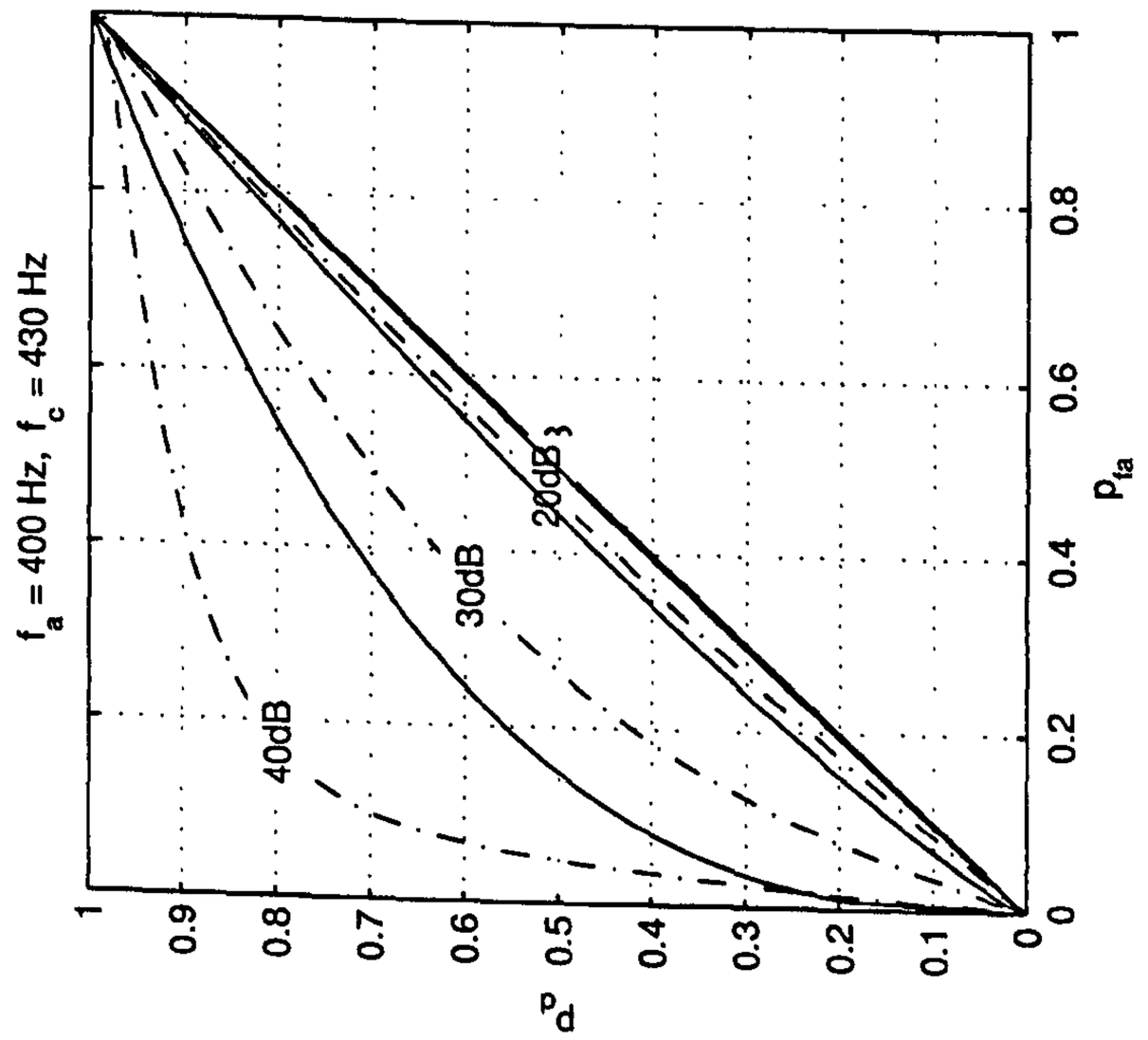


Figure 5.3: ROC curves for the squared-envelope detector (solid lines) and the continuous interval detector (dash-dotted lines). To avoid clutter, only the interval detector curves are labelled with the SNR. Note that the high-SNR ROC curves for the squared-envelope detector are located in the upper left-most portion of the plot for $f_c = 400$ Hz and $f_c = 410$ Hz.



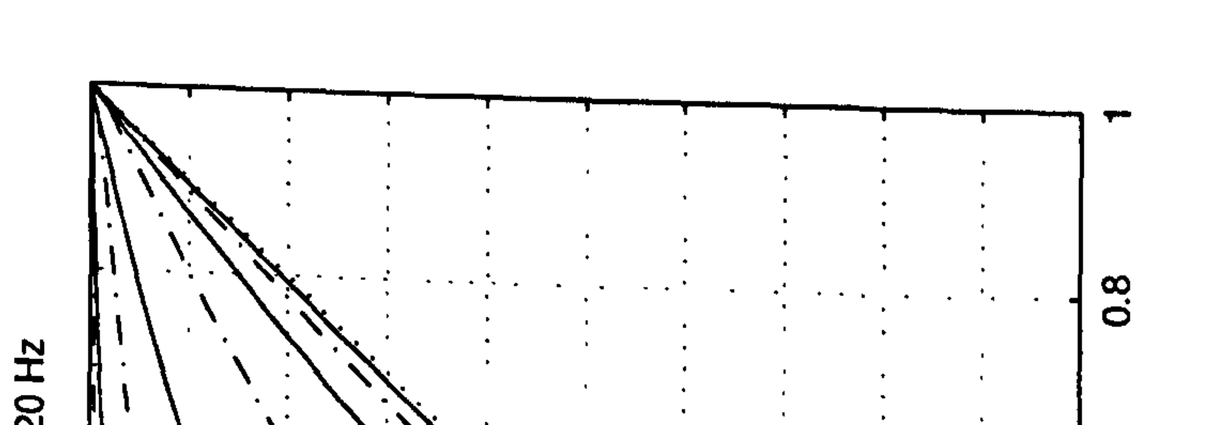


Figure 5.3: ... continued.

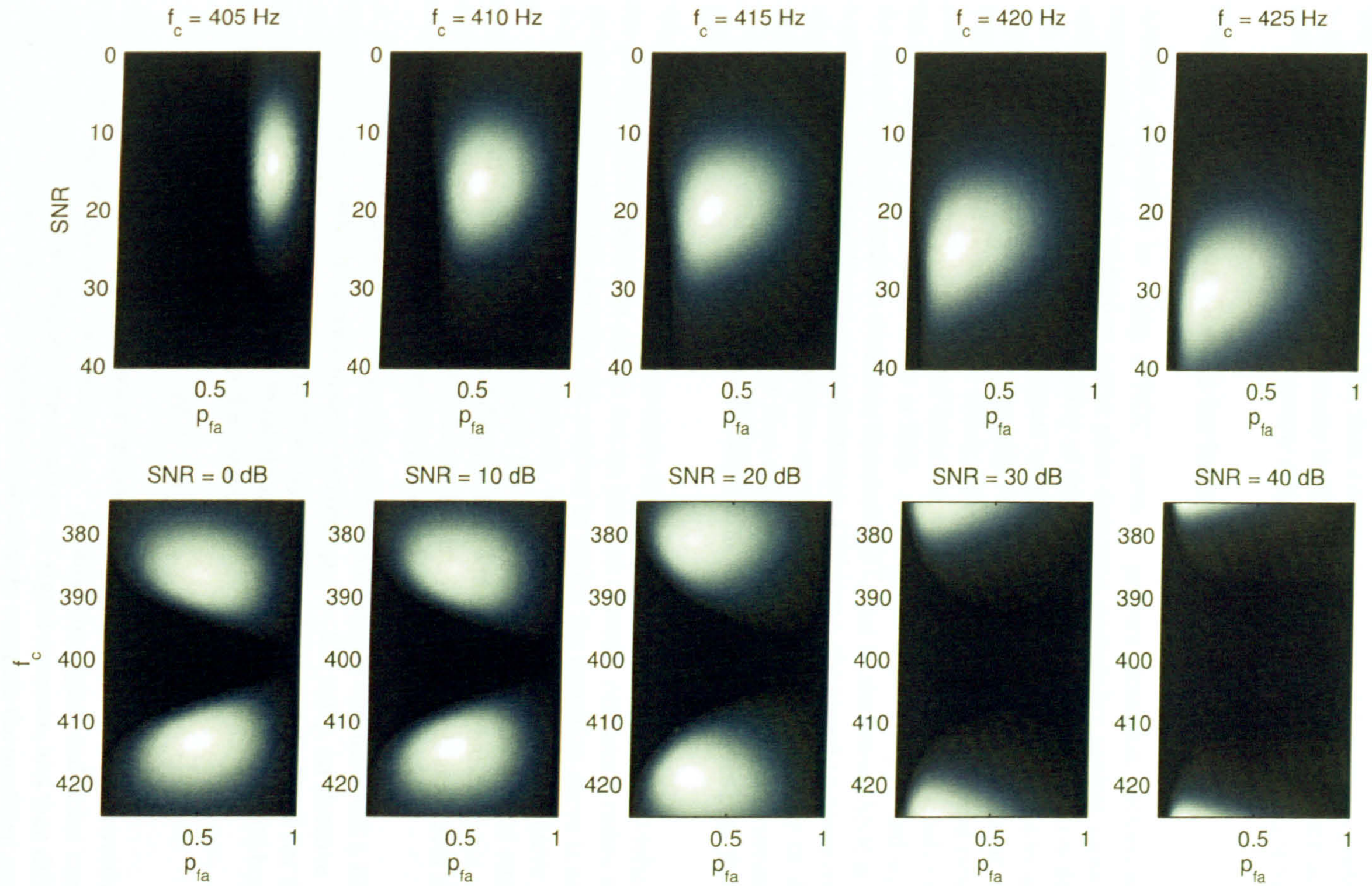


Figure 5.4: Regions of superior performance for the continuous interval detector. See discussion in Section 5.1.3.

representation of $p'_d(p_{fa}, f_c)$, with the SNR held fixed at five different values. This kind of image allows one to choose which detector to employ in specific circumstances. For instance, if the probability of false alarm is to be fixed at 0.25, the SNR is 10 dB and the signal frequency is 415 Hz, then a continuous interval detector is preferable. It is also seen that if the signal frequency is 405 Hz, then the continuous interval detector is inadmissible for all but a very narrow range of parameters in which p_{fa} is very high.

5.1.4 Producing Transition Curves for Interval Detectors

Another alternative to using a ROC curve to present detection performance is a *transition curve*. A transition curve plots detection probability against signal excess, whilst maintaining a fixed probability of false alarm. Signal excess (SE) is defined as the difference between the true signal level and the signal level required to achieve a nominal probability of detection (usually 50%), with respect to a particular choice of detection threshold (Dawe, 1997; Urick, 1976). A transition curve would typically be used as follows. The probability of false alarm is set to some suitably small value, and the probability of detection is set at 50%. A particular choice of p_{fa} and p_d dictates the detection index and the detection threshold, DT, that is, the lowest SNR at which it is possible to achieve the preset performance. Of course, the signal-to-noise conditions vary greatly at sea, and the transition curve shows how the probability of detection improves when the SNR is higher than expected (SE $>$ 0 dB), or deterioriates when lower than expected (SE $<$ 0 dB). By definition, all transition curves pass through the point $(0, 50\%)$.

Whilst one can generate a transition curve for a power-based detector with relative ease, the production of transition curves for an interval detector requires a rather awkward series of computations. Ideally, the manner in which the transition curve is employed would suggest how it is to be computed: first, the SNR required to achieve p_{fa} and $p_d = 0.5$ is calculated (i.e., DT); next, the decision regions are identified and held in place; and lastly, the probability of detection is plotted as the signal excess is varied. It is difficult to carry out the first step because there is no practical, analytical means of mapping a (p_{fa}, p_d) pair to a detection threshold.

A more circuitous solution is to fix the detection threshold and then find a likelihood threshold, $\eta_{50\%}$, which approximately satisfies $p_d = 0.5$ using an iterative scheme, such as interval bisection. The probability of false alarm is calculated for the $\eta_{50\%}$ parameter and subsequently used to label the transition curve. The decision boundaries are identified by zero crossings in the function

$$p_{I_c}(i_c \mid H_1) - \eta_{50\%}\, p_{I_c}(i_c \mid H_0).$$

An upward crossing reveals the lower decision boundary, i_{c0}, and a downward crossing reveals the upper decision boundary, i_{c1}. Finally, the signal-and-noise cumulative distribution function is computed for a variety of signal excesses, and then differenced at the upper and lower decision boundaries in order to compute the modified probability of detection, i.e.,

$$p_d(\mathsf{SE}) = P(I_c \leq i_{c1}) - P(I_c \leq i_{c0}).$$

Figure 5.5 presents a collection of four transition curves generated in this way.

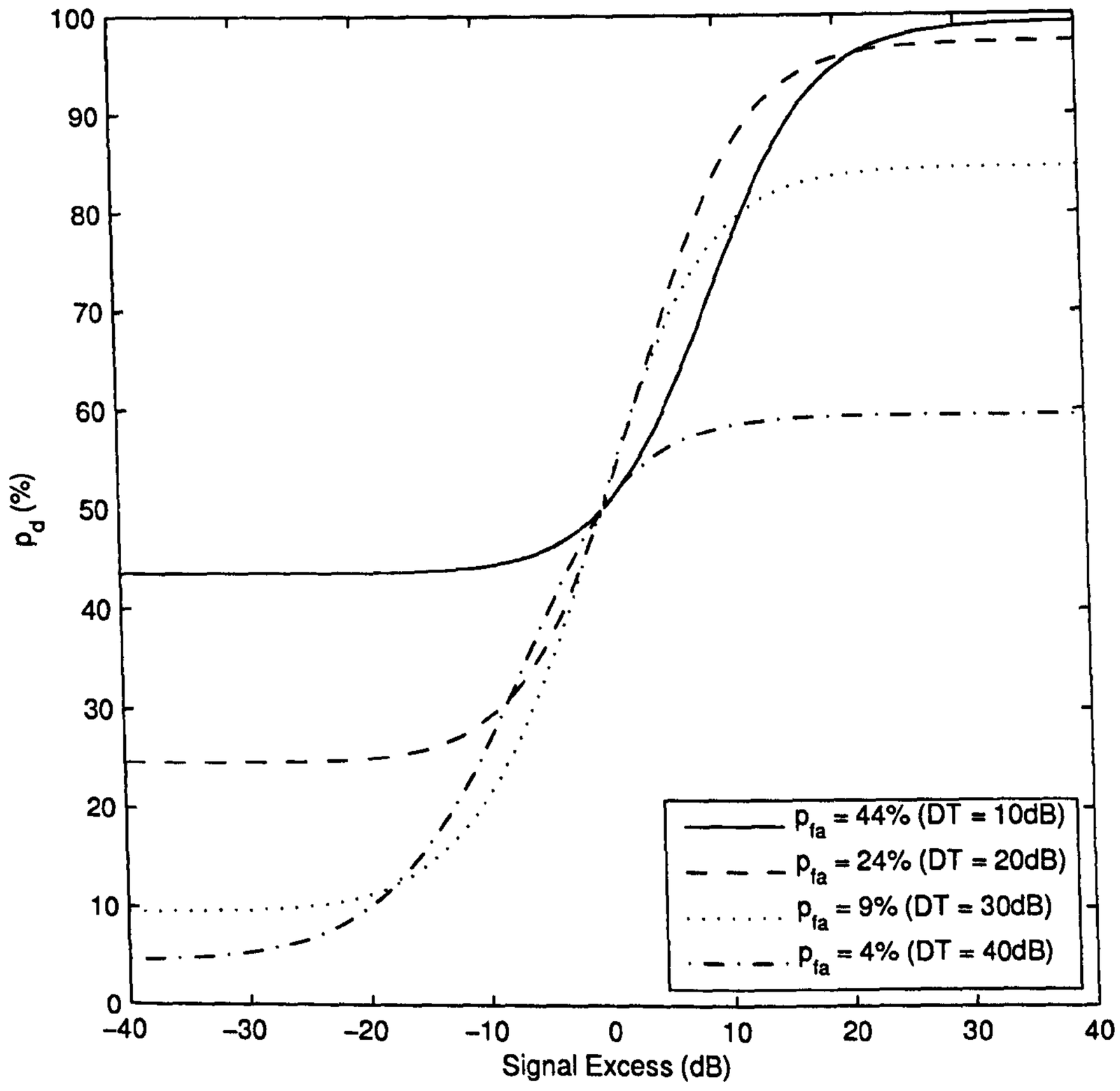


Figure 5.5: Transition curves for a continuous interval detector, calibrated to give a 4%, 9%, 24% and 44% probability of false alarm. In each case, when the signal excess is 0 dB, a 50% detection probability is assured. The legend lists the detection thresholds from which the probability of false alarms were derived at $\mathsf{SE} = 0$.

5.2 Detection of a Sinusoid

In the derivations and detection studies reported earlier, we assumed that the target signal and background noise were zero-mean stationary Gaussian processes. This assumption was made largely to ease the derivation of the conditional probability density functions upon which the detectors rely, but at the expense of restricting the application of the detector to just one class of signal model.

The passive sonar literature is chiefly concerned with the detection of *sinusoidal* target signals against a Gaussian noise background, making the detection of a sinusoid in noise the most relevant line of enquiry when first venturing beyond the stationary Gaussian targets considered so far. In this section, the signal to be detected is a random process of the form

$$A\cos(\omega_c t + \theta), \tag{5.14}$$

where A and ω_c are constants that set the amplitude and frequency, respectively, and $\theta \sim \text{Uniform}\{-\pi, \pi\}$. Random processes of the form given in (5.14) are referred to hereafter as *randomly-phased sinusoids*. The goal, then, is to detect a randomly-phased sinusoid in the presence of additive stationary Gaussian noise using a zero crossing interval.

A pure, noise-free sinusoid with radial frequency ω only generates intervals equal to its half-period, so the continuous interval p.d.f. in this special case is self-evidently

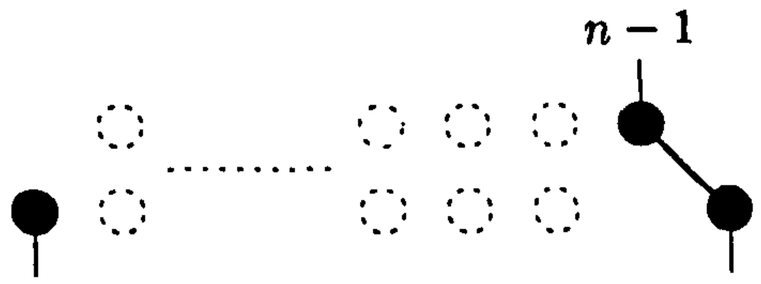
$$p_{I_c}(i_c) = \delta\left(i_c - \frac{\pi}{\omega_c}\right). \tag{5.15}$$

Deriving the interval p.d.f. for a sinusoid mixed with noise presents a greater challenge however. If the sinusoid has uniformly-random phase, then the process is stationary but non-Gaussian; if the phase of the sinusoid is fixed, then the process is Gaussian but non-stationary: either case violates one of the requirements set out above (Whalen, 1971). As our main focus is on the first case, we must either: i) re-derive the interval distribution for a sinusoid in noise from first principles; ii) disregard the Gaussian assumption and proceed as though the target process *were* Gaussian, incurring some penalty from the mismatch between the model and real data; or iii) adapt the method or process in such a way that the two become mutually compatible. The three approaches that are presented next examine each of these options in turn.

5.2.1 A Derivation Specific to a Sinusoid in Noise

Earlier, we arrived at the interval density function by considering the probability of sign changes in a random process. The probability of a particular pattern of sign changes was evaluated by integrating the joint density function for the samples of the process in an orthant region. We shall take the same approach in this section, only this time the distribution governing a set of samples will be non-Gaussian. Other approaches to determining the probability of an interval, i.e., those based on something other than the sign changes of a stationary random process, are not considered here (but review Section 1.2.2 above).

The cumulative distribution function governing the zero crossing interval for a general wide-sense stationary process was determined in Section 4.3.2 (with intervals restricted to $k_0 < i < 2k_0$ for some k_0) in terms of two orthant probabilities:



$$P(I \leq k) = \begin{cases} \dfrac{2P(x_n < 0, x_{n-1} \geq 0, x_{n-k-1} < 0)}{1 - 2P(x_n \geq 0, x_{n-1} \geq 0)} & k_0 < k < 2k_0 \\ 0 & k \leq k_0 \\ 1 & k \geq 2k_0. \end{cases} \tag{5.16}$$

Note that the formulation in (5.16) also assumes that a sequence of sign changes and its negation (i.e., the same sequence "upside-down") are equiprobable. Tackling this problem for a sinusoid in noise requires first obtaining an expression for the joint p.d.f. for two and three samples then integrating the appropriate orthant. We shall start by obtaining the distribution for a single sample in the random process.

The received process X is an additive mixture of a sine wave S and Gaussian noise samples V. The distribution governing the samples of a clean sinusoid with amplitude $A > 0$ can be found by considering a single cosine period, normalised between $-\pi$ and π. Consulting the sketch in Figure 5.6, the cumulative distribution function of S is seen to be (Bendat and Piersol, 2000)

$$P(S \leq s; A) = \begin{cases} 1 - \dfrac{1}{\pi} \arccos\left(\dfrac{s}{A}\right) & -A < s < A \\ 0 & s \leq -A \\ 1 & s \geq A. \end{cases} \tag{5.17}$$

The probability density function governing S is the derivative of (5.17).

$$p_S(s; A) = \frac{dP(S \leq s; A)}{ds} = \begin{cases} \dfrac{1}{\pi\sqrt{A^2 - s^2}} & -A < s < A \\ 0 & \text{otherwise.} \end{cases} \tag{5.18}$$

The individual samples of the noise process V are governed by a Gaussian distribution with variance σ^2, i.e.,

$$p_V(v) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{v^2}{-2\sigma^2}\right). \tag{5.19}$$

The probability density function governing the sum of two independent random variables is found by convolving the p.d.f.s of the individual variables (Peebles, 1993).

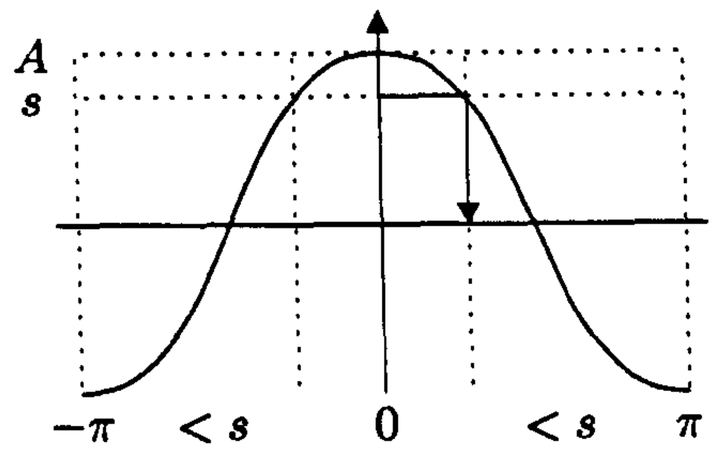


Figure 5.6: A single cosine period.

As the sinusoid and noise samples are independent, it follows that the p.d.f. of the noise-added sinusoid is given by the convolution integral

$$p_X(x;A) = \frac{1}{\sqrt{2\pi^3\sigma^2}} \int_{\mathcal{A}} \frac{\exp(-(x-\xi)^2/(2\sigma^2))}{\sqrt{A^2-\xi^2}} d\xi, \ \mathcal{A} = (-A, A). \qquad (5.20)$$

An alternative approach to finding the probability density function, outlined by Whalen (1971), is determined by means of characteristic functions and takes the form of two nested infinite series[1]. For an alternative expression, see Bendat and Piersol (2000).

For our purposes, it seems that neither the integral solution given in (5.20) nor the infinite series provided by Whalen is a workable option, especially considering that up to now we have only dealt with individual samples. Looking ahead to the joint probability density for two or three samples, the author surmises that it is possible to write down an expression for these p.d.f.s in integral form[2], but integration of the resulting expression for the p.d.f. in an orthant region, by all but numerical methods, lies firmly beyond reach.

5.2.2 Treating a Sinusoid in Noise as a Gaussian Process

As directly deriving the interval distribution fails, the second option is to employ the interval density function derived for Gaussian processes, ignore the fact that a sinusoid in noise is non-Gaussian, and tolerate a certain loss in performance. If the analytical and empirical interval p.d.f.s appear to compare favourably, then some formal work can be undertaken to establish bounds on the model error; if not, the method will be

[1] See Whalen (1971), page 100.

[2] For instance, the solution to the joint p.d.f. for two samples is

$$p_{X_n X_{n+k}}(x_n, x_{n+k}) = \frac{1}{4\pi^2|\Sigma|^{1/2}} \int_{\mathcal{A}} (A^2-\xi^2)^{-1/2} \left[\exp \frac{\mathbf{s}_+^T \Sigma^{-1} \mathbf{s}_+}{-2} + \exp \frac{\mathbf{s}_-^T \Sigma^{-1} \mathbf{s}_-}{-2} \right] d\xi$$

where, given the covariance matrix Σ, sampling interval Δt and sinusoid frequency ω_c,

$$\mathbf{s}_\pm = \begin{bmatrix} \xi \\ \xi \cos(\omega_c n \Delta t) \mp \sqrt{1-\xi^2} \sin(\omega_c n \Delta t)) \end{bmatrix}, \ \mathcal{A} = (-A, A).$$

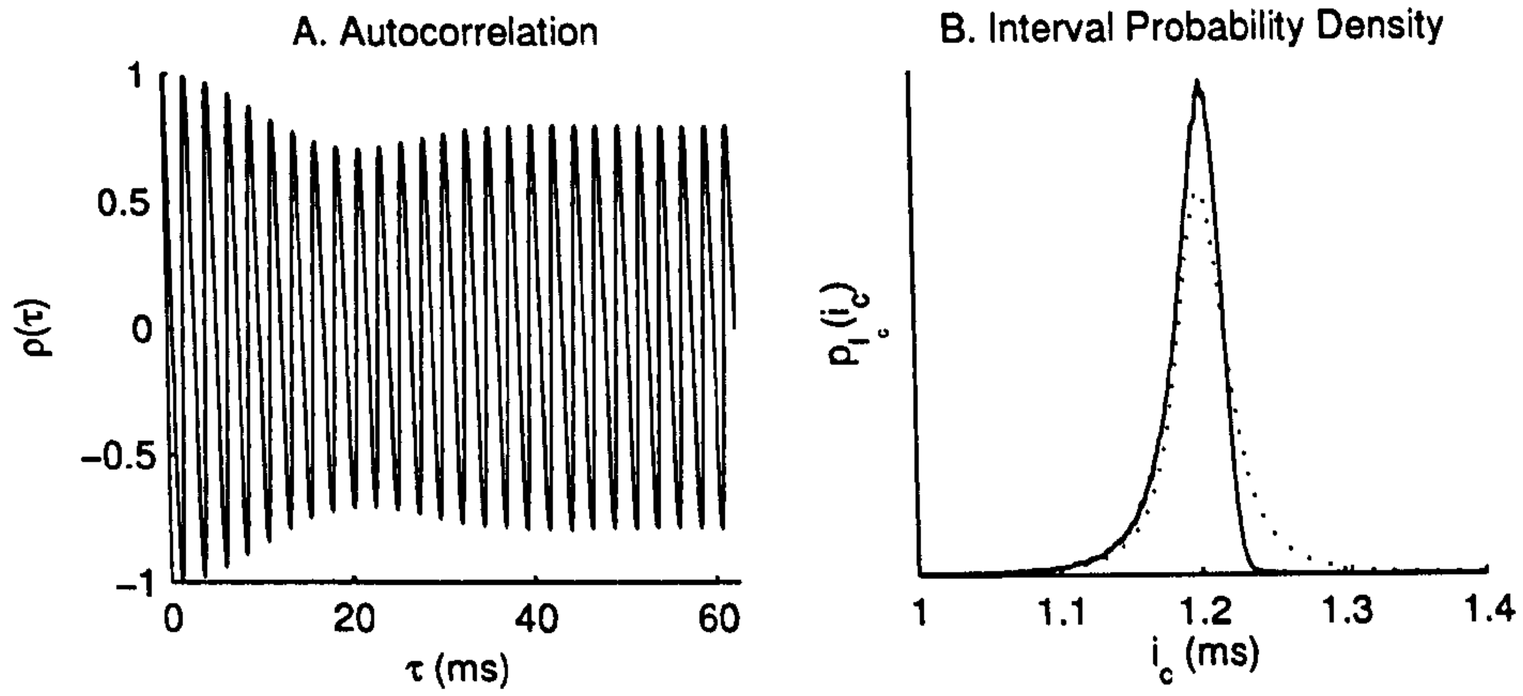


Figure 5.7: Autocorrelation function (A) found empirically (solid line) and analytically (obscured dotted line); interval probability density function (B) obtained empirically (solid line) and analytically, using a limiting approach (dotted line).

deemed inadmissable. The first step is to decide upon a Gaussian process that is a suitable candidate for replacing a sinusoid.

A pure sinusoid of infinite duration manifests in the frequency domain as two Dirac delta functions at a positive and negative frequency. An impulse of this form may be obtained as the limit of a Gaussian function as its width tends to zero, whilst its area is held constant (Peebles, 1993). This suggests modelling a sinusoidal process as a concentration of energy at a single frequency, or as noise passed through a filter whose width is vanishingly small. If we convolve white Gaussian noise with the familiar impulse response

$$h_s(t) = \exp(-2(C(t-\mu))^2)\cos\omega_c t \tag{5.21}$$

and allow $C \to 0$, then the autocorrelation function becomes

$$\rho(\tau) = \cos\omega_c\tau \tag{5.22}$$

in the limit, and for the c.d.f. for the interval distribution from (4.69) we get

$$P(I_c \leq \tau) = \begin{cases} \frac{1}{2}\left(1 - \frac{\sin\omega_c i_c}{|\sin\omega_c i_c|}\right) & \tau_0 < i_c < 2\tau_0 \\ 0 & i_c \leq \tau_0 \\ 1 & i_c \geq 2\tau_0. \end{cases} \tag{5.23}$$

This cumulative distribution (5.23) constitutes a step function from 0 to 1, in which the discontinuity occurs at $i_c = \pi/\omega_c$. This is of course consistent with the intervals of a pure sinusoid with frequency $f_c = 2\pi\omega_c$. Unfortunately, a sinusoid added to band-pass Gaussian noise cannot be modelled in this naïve fashion. The equations governing the changes of sign in a Gaussian process, derived earlier, hold when the filter is extremely narrow but become degenerate in the limit.

Figure 5.7 compares the interval distribution obtained by modelling a sinusoid as an infinitely-narrowband Gaussian process with a histogram formed from the intervals of

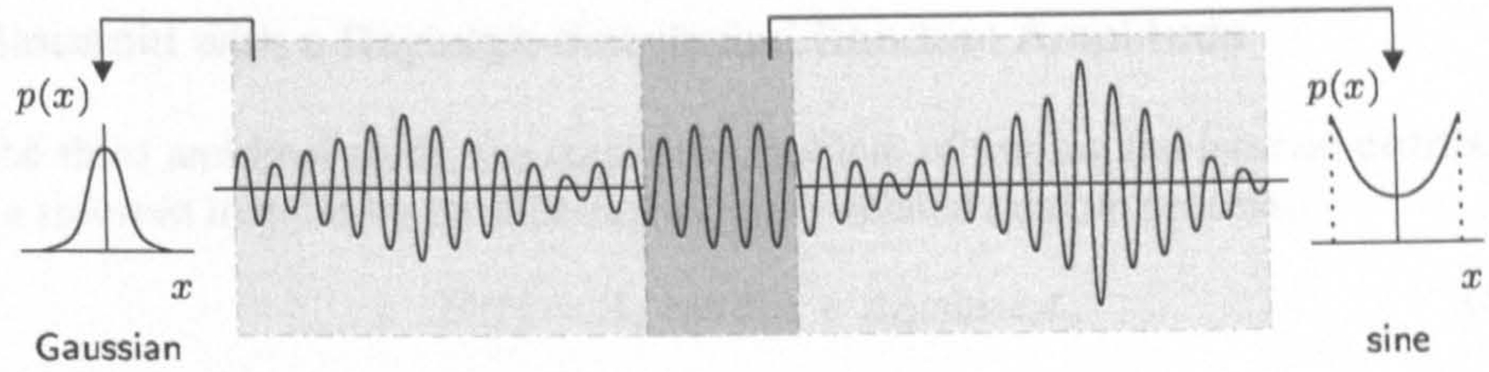


Figure 5.8: Illustrative probability density functions for the samples of a narrowband process measured over a short (dark grey) and long (light grey) time scale.

a random signal. In this case, the random signal is a 420 Hz sinusoid added to white noise with a narrowband SNR of 30 dB, filtered with a band-pass filter whose impulse response has the MGMM description

$$\left\langle \frac{1}{2}, 40, 0, +2\pi \cdot 400, 0 \right\rangle + \left\langle \frac{1}{2}, 40, 0, -2\pi \cdot 400, 0 \right\rangle . \tag{5.24}$$

The procedure for finding a time-invariant autocorrelation function generalises to any wide-sense stationary process, including those whose samples are non-Gaussian distributed, so the agreement between the analytical and empirical autocorrelation functions in Figure 5.7A is expected. On the other hand, Figure 5.7B reveals a marked difference between the analytical and observed interval probability density functions; in particular, the true p.d.f. exhibits a greater degree of asymmetry than the model. These differences must be attributed to the non-Gaussian distribution of the signal samples, because the autocorrelation functions are identical, and the interval p.d.f. is a function only of the autocorrelation. Maximum likelihood detectors rely upon a good match between the model conditional distributions and the true distribution of the data to perform well, so the results in Figure 5.7B clearly rule out the approach attempted in this section.

Consideration of the envelope of a Gaussian process offers a further insight into why the limiting approach fails. A narrowband Gaussian process can be envisaged as a sinusoid with a slowly fluctuating envelope and phase: in the short term, the samples of the process show a sinusoidal distribution, but in the long term, the samples are Gaussian-distributed (Figure 5.8). If such a signal is mixed with a band-pass process, then, during the wax of the signal envelope, the mixture and its intervals will be dominated by the signal; conversely, during the wane of the signal envelope, the mixture will be dominated by noise, and the intervals will exhibit higher variance. If the signal bandwidth is reduced, the envelope varies at a lower frequency, i.e., the waxes and wanes drift further apart, and it becomes necessary to observe a longer window of the signal before the empirical interval distribution and the true population distribution converge. If the signal bandwidth is infinitesimal—our present strategy for modelling a sinusoid—then the signal envelope is constant, and the samples of the process never converge to a Gaussian distribution.

5.2.3 A Sinusoid with a Rayleigh-distributed Random Amplitude

In the third and final study, we tackle the problem of finding the interval distribution for a sinusoid in noise by considering a closely-related random process,

$$G(t) = A_I \cos \omega_c t + A_Q \sin \omega_c t, \tag{5.25}$$

where A_I and A_Q are independent Gaussian random variables with zero mean and equal variance v^2, and ω_c is a constant parameter controlling frequency. This type of process is both Gaussian *and* wide-sense stationary, and is mentioned in connection with the zero crossing rate by Kedem (1986). For a given time t, the cosine and sine terms are constant, so $G(t)$ is the weighted sum of two independent Gaussian variables. Also, the mean of the process G is a constant,

$$\begin{aligned} E\{G\} &= E\{A_I \cos \omega_c t + A_Q \sin \omega_c t\} && (5.26)\\ &= E\{A_I\} \cos \omega_c t + E\{A_Q\} \sin \omega_c t && (5.27)\\ &= 0, && (5.28) \end{aligned}$$

and its autocovariance is time-invariant (i.e., does not depend on t),

$$\begin{aligned} E\{G_t G_{t-\tau}\} &= E\{(A_I \cos \omega_c t + A_Q \sin \omega_c t) \times \\ &\quad (A_I \cos \omega_c (t-\tau)) + A_Q \sin \omega_c (t-\tau))\} && (5.29)\\ &= E\{A_I^2\} \cos(\omega_c t) \cos(\omega_c (t-\tau)) \\ &\quad + E\{A_Q^2\} \sin(\omega_c t) \sin(\omega_c (t-\tau)) && (5.30)\\ &= v^2 \cos \omega_c \tau, && (5.31) \end{aligned}$$

so the process is wide-sense stationary. Because this random process fulfils these two key requirements, its autocorrelation function (5.31) can be used to find the distribution of its zero crossing intervals, by application of (4.70).

The Rayleigh Density Function

The random process described by (5.25) closely resembles the type of process we are aiming to model—a randomly-phased sinusoid—but it is not quite the same. The randomly-phased sinusoid described in (5.14) can be expanded into

$$A \cos(\omega_c t + \theta) = A \sin \theta \cos \omega_c t + A \cos \theta \sin \omega_c t, \tag{5.32}$$

which is identical in format to (5.25), setting $A_I = A \sin \theta$ and $A_Q = A \cos \theta$. The joint distribution of A_I and A_Q is bivariate Gaussian with p.d.f.

$$p_{A_I A_Q}(a_I, a_Q) = \frac{1}{2\pi v^2} \exp \frac{{a_I}^2 + {a_Q}^2}{-2v^2}. \tag{5.33}$$

Transforming (5.33) into polar coordinates, we obtain the joint probability density for A and θ,

$$p_{A\Theta}(a, \theta) = \frac{a}{2\pi v^2} \exp \frac{a^2}{-2v^2}, \; a > 0, \tag{5.34}$$

with the marginal densities

$$p_A(a) = \frac{a}{\upsilon^2} \exp \frac{a^2}{-2\upsilon^2},\ a > 0 \tag{5.35}$$

$$p_\Theta(\theta) = \frac{1}{2\pi}. \tag{5.36}$$

The random process in (5.25) is equivalent to a sinusoid with uniform-random phase, as the marginal density (5.36) makes plain. However, the amplitude of the sinusoid is now a *random variable*, governed by the well-known *Rayleigh distribution* (Peebles, 1993). The Rayleigh distribution has a single parameter, υ^2, and to indicate that A is a Rayleigh variable, the following notation is used.

$$A \sim \text{Rayleigh}\{\upsilon^2\}\ (\text{or R}\{\upsilon^2\})$$

$$p_A(a;\upsilon^2) = \frac{a}{\upsilon^2} \exp \frac{a^2}{-2\upsilon^2},\ a > 0.$$

We shall now stand back for a moment and review the various random processes we have encountered. Our initial aim was to discover the distribution governing the intervals of a sinusoid with fixed amplitude, A, and unknown phase, θ, in Gaussian noise. However, because that process is non-Gaussian, it is hard to reconcile with the theoretical framework laid out already, which strongly rests on the assumption of Gaussian distribution. In this section, another sinusoidal random process has been introduced, $G(t)$, whose properties include stationarity, Gaussian distribution and uniform random phase. Now, however, the amplitude of the sinusoid is random. Note that here 'random' does not imply that the envelope of a sample function varies randomly with time, rather that it is fixed at a random value from the outset (drawn from a Rayleigh distribution), and we would like it to be fixed at some constant.

5.2.4 A Sinusoid with a Constant Amplitude

Our goal is to find the probability density function that governs the intervals of a randomly-phased sinusoid with amplitude A, mixed with a Gaussian noise process. We do not yet have a means of evaluating this p.d.f. if A is fixed. However, we concluded above that if A is drawn from a Rayleigh distribution, the probability density function governing I_c is known:

$$p_{I_c}(i_c \mid A \sim \text{R}\{\upsilon^2\}) = \int_0^\infty p_{I_c}(i_c \mid A = a) p_A(a;\upsilon^2) da. \tag{5.37}$$

The expression (5.37) corresponds to the probability density function governing the intervals of a sinusoid in noise, whose amplitude is initially chosen according to a Rayleigh distribution with parameter υ^2. Next, let us for a moment imagine that it is possible to find a function $w(\upsilon)$ which satisfies

$$\int_0^\infty w(\upsilon) p_A(a;\upsilon^2) d\upsilon = \delta(a - a_0). \tag{5.38}$$

A constant amplitude a_0, and a random amplitude whose p.d.f., $\delta(a-a_0)$, concentrates all the probability mass upon a single value a_0, are equivalent. Hence,

$$p_{I_c}(i_c \mid A = a_0) = \int_0^\infty p_{I_c}(i_c \mid A = a)\delta(a - a_0)da \tag{5.39}$$

$$= \int_0^\infty p_{I_c}(i_c \mid A = a)\left[\int_0^\infty w(v)p_A(a; v^2)dv\right]da \tag{5.40}$$

$$= \int_0^\infty w(v)\left[\int_0^\infty p_{I_c}(i_c \mid A = a)p_A(a; v^2)da\right]dv. \tag{5.41}$$

The bracketed expression in (5.41) is the same as the right-hand side of (5.37), for which we possess a solution; all that remains is to find $w(v)$.

It is not clear whether it is possible to obtain a continuous function $w(v)$ that satisfies (5.38). The practical alternative pursued here is to solve for an approximation using sampled Rayleigh density functions,

$$p_A[a; v^2] = \frac{a}{v^2}\exp\frac{a^2}{-2v^2}, \tag{5.42}$$

and weighting function, $w[v]$, in which a and v are discrete. In the following analysis, the target amplitude has been set, without loss of generality, to $a_0 = 1$, and a and v have been chosen by hand to assume the values

$$a_i = 0.01i,\ 1 \leq i \leq 200 \tag{5.43}$$

$$v_j = 0.02j,\ 1 \leq j \leq 25. \tag{5.44}$$

A discrete approximation of (5.38) may be expressed in matrix form:

$$R\mathbf{w} = \mathbf{t}, \tag{5.45}$$

where R is a 200×25 matrix, whose columns are populated with Rayleigh density functions,

$$[R_{i,j}] = \frac{a_i}{v_j^2}\exp\frac{a_i^2}{-2v_j^2}, \tag{5.46}$$

$\mathbf{w}$ is the 25×1 weight vector that we wish to find, and $\mathbf{t}$ is the 200×1 "target" vector, whose entries represent the amplitude density function we wish to approximate, i.e.,

$$\mathbf{t}_i = \begin{cases} 100 & \text{if } a_i = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{5.47}$$

Generally, there does not exist a vector $\mathbf{w}$ that precisely satisfies (5.45); instead a standard least-mean squares (LMS) technique is used to minimise a cost function expressing the squared-error between the weighted sum of Rayleigh p.d.f.s and the target p.d.f., i.e.,

$$J(\mathbf{w}) = (R\mathbf{w} - \mathbf{t})^2. \tag{5.48}$$

The procedure for choosing $\mathbf{w}$ to minimise $J(\mathbf{w})$ is ubiquitous and is described in, e.g., Bishop and Hinton (1995). As the steps required are few, we will reproduce them here

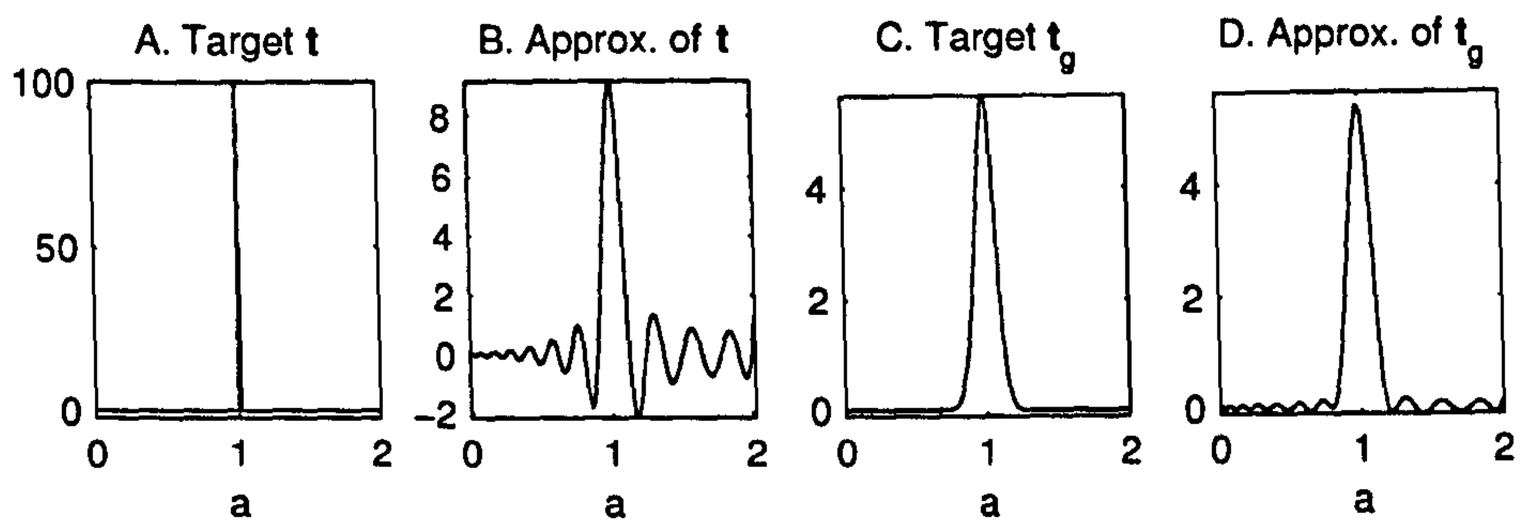


Figure 5.9: A. a spike at a_0; B. an approximation of the spike (by a sum of twenty-five Rayleigh p.d.f.s); C. a narrow pulse at a_0; D. an approximation of the pulse.

for completeness. The cost function $J(\mathbf{w})$ is convex, so the cost-minimum $\hat{\mathbf{w}}$ occurs at its (unique) turning point, i.e.,

$$\nabla_w\{J\}(\hat{\mathbf{w}}) = 0. \tag{5.49}$$

So first differentiate with respect to $\mathbf{w}$,

$$\begin{aligned}
\nabla_w\{J(\mathbf{w})\} &= \nabla_w\{(R\mathbf{w}-\mathbf{t})^2\} && (5.50)\\
&= \nabla_w\{(R\mathbf{w}-\mathbf{t})^T(R\mathbf{w}-\mathbf{t})\} && (5.51)\\
&= \nabla_w\{\mathbf{w}^T R^T R\mathbf{w} + \mathbf{t}^T\mathbf{t} - 2\mathbf{w}^T R^T\mathbf{t}\} && (5.52)\\
&= 2R^T R\mathbf{w} - 2R^T\mathbf{t}. && (5.53)
\end{aligned}$$

then equate the solution with zero to find the LMS optimum $\hat{\mathbf{w}}$:

$$\begin{aligned}
R^T R\hat{\mathbf{w}} &= R^T\mathbf{t} && (5.54)\\
\hat{\mathbf{w}} &= (R^T R)^{-1}R^T\mathbf{t}. && (5.55)
\end{aligned}$$

The quantity $(R^T R)^{-1}R^T$ is the pseudoinverse of R (Bishop and Hinton, 1995), often denoted $R^\dagger$ or R^+, and is computed efficiently by the MATLAB function `pinv`.

The target p.d.f., $\mathbf{t}$, is plotted in Figure 5.9A, accompanied by the approximation $R\hat{\mathbf{w}}$ in Figure 5.9B. The approximated p.d.f. succeeds in concentrating the probability mass around $a_0 = 1$, but there are also a few unwanted oscillations on either side, making $R\hat{\mathbf{w}}$ inadmissable, as negative values relate negative probability mass.

This "ringing" effect is a consequence of trying to model the sharp discontinuity in $\mathbf{t}$ as the sum of a finite number of smooth functions, similar to that which arises when approximating a square pulse with a low-order Fourier series. One solution, in either scenario, is to replace the sharp function with a smoother version; in this case, we employ a Gaussian pulse centred on $a_0 = 1$ instead of a spike:

$$[\mathbf{t}_g]_i = \frac{1}{\sqrt{2\pi q}}\exp\frac{(0.01i-1)^2}{-2q},\ 1 \le i \le 200, \tag{5.56}$$

where q controls the width of the pulse and is set at 0.01. The elements of the vector $\mathbf{t}_g$ are plotted in Figure 5.9C. (One informal interpretation of this solution says, "If there

is any departure from zero, we would rather it were in the vicinity of a_0.") A new set of weights is then computed using

$$\hat{\mathbf{w}}_g = (R^T R)^{-1} R^T \mathbf{t}_g, \tag{5.57}$$

leading to the approximation $R\hat{\mathbf{w}}_g$, which is shown in Figure 5.9D. From a visual inspection, $R\hat{\mathbf{w}}_g$ is a close approximation of $\mathbf{t}_g$, and negative probability mass is no longer a serious problem. (A narrower pulse is obtained by setting $q = 0.005$, but a small amount of ringing reappears.) By using different values for a_i and v_j, it may be possible to find a better approximation; however, the particular solution presented above will suffice, as its application in the next section demonstrates.

5.2.5 The Interval Distribution for a Sinusoid in Noise

Combining the results up to this point, the probability density function for the intervals of a sinusoid with unit amplitude in noise is seen to have the approximation[1]

$$p_{I_c}(i_c \mid A = 1) \approx \sum_{j=1}^{25} [\hat{\mathbf{w}}_g]_j \int_0^\infty p_{I_c}(i_c \mid A = a) p_A(a; v_j^2) da \tag{5.58}$$

$$= \sum_{j=1}^{25} [\hat{\mathbf{w}}_g]_j \cdot \underbrace{p_{I_c}(i_c \mid A \sim \mathrm{R}\{v_j^2\})}, \tag{5.59}$$

where $\mathbf{w}_g$ is computed according to the previous section, $p_A(\cdot)$ is the Rayleigh p.d.f., and the braced expression is computed using MGMMs (and incorporates the details of the noise distribution). This procedure can be extended to a sinusoid with general amplitude, a_0, by appropriately dilating the Rayleigh p.d.f.s:

$$p_{I_c}(i_c \mid A = a_0) \approx \sum_{j=1}^{25} [\hat{\mathbf{w}}_g]_j \cdot p_{I_c}(i_c \mid A \sim \mathrm{R}\{a_0^2 v_j^2\}). \tag{5.60}$$

Note that the coefficients $\hat{\mathbf{w}}_g$ are unchanged and therefore need only be calculated once.

Section 5.2.2 described an attempt to derive an interval probability density function by ignoring the non-Gaussian distribution of the samples. A comparison of the p.d.f. with an interval histogram in Figure 5.9 demonstrated the failure of this approach. In this section, we carry out a similar evaluation of the "Rayleigh-sum" technique that we have just described. The random process is once again a 420 Hz tone mixed with white noise at 30 dB narrowband SNR, and the impulse response of the analysis filter is specified in (5.24). Figure 5.10A plots three interval probability density functions: a histogram formed by measuring the intervals of a synthesised process, which represents the ground truth (solid line); the analytical approximation computed using the Rayleigh-sum method (dash-dotted line); and the analytical approximation based on the naïve assumption of Gaussian distribution (dotted line). The analytical

[1] This particular spacing of twenty-five Rayleigh p.d.f.s was the result of trial-and-error. Including more curves appeared to confer little benefit; reducing the number of curves led to degradation.

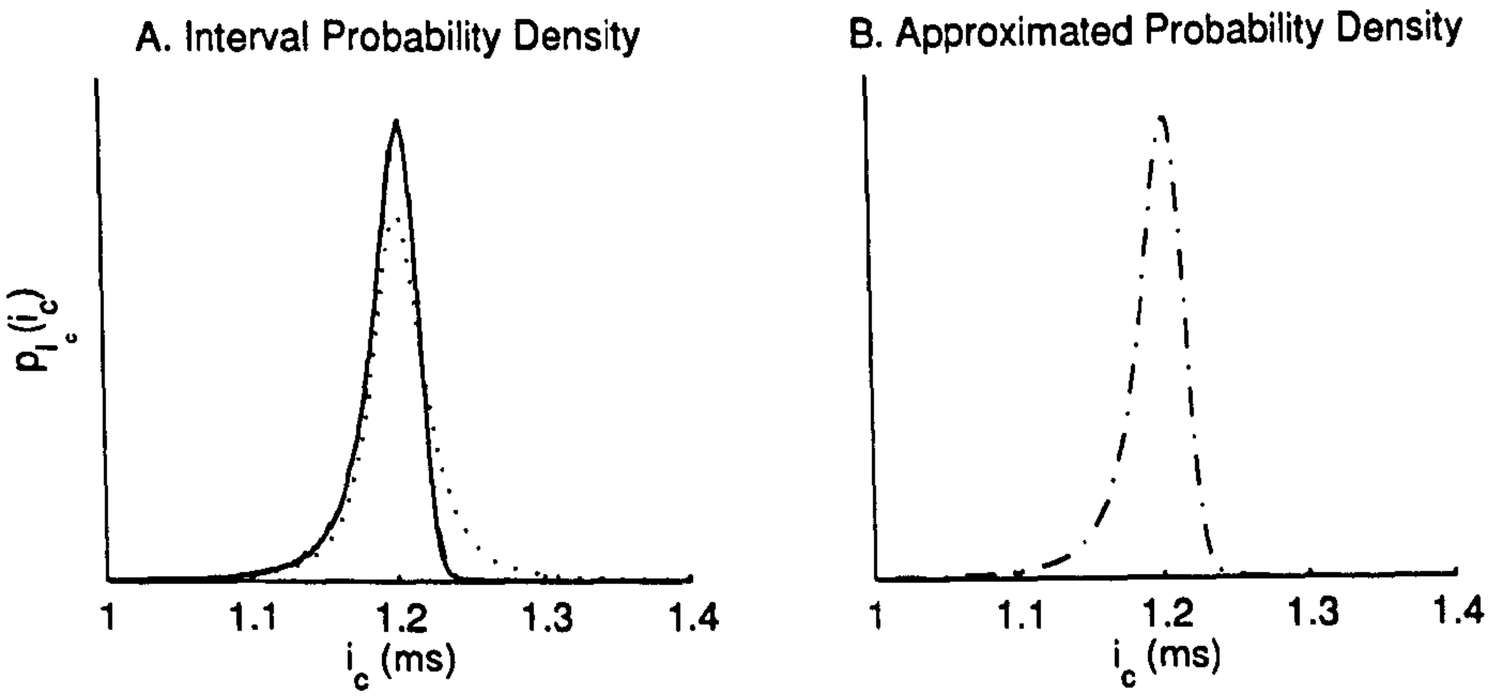


Figure 5.10: Interval probability density function for a sinusoid in noise. A) empirical (solid), analytical using Rayleigh-sum approach (dash-dotted / obscured) and a limiting approach (dotted); B) Rayleigh-sum analytical p.d.f. reproduced from (A).

p.d.f. produced using the Rayleigh-sum method is virtually indistinguishable from the histogram, and constitutes a significant improvement over the naïve p.d.f., most notably in its ability to reproduce the asymmetrical character of the distribution. (The analytical p.d.f. is difficult to discern, so it is plotted again in Figure 5.10B.)

5.2.6 An Interval-based Sinusoid Detector

The interval-based sinusoid detector is a maximum-likelihood detector, and it operates according to the minimum error decision rule:

$$\text{choose } H_1 \text{ iff } \frac{p_{I_c}(i_c \mid H_1)}{p_{I_c}(i_c \mid H_0)} > 1, \text{ otherwise choose } H_0.$$

The only difference between this detector and the continuous interval detector is that the conditional probability density function $p_{I_c}(i_c \mid H_1)$ in the decision rule uses (5.60) instead of (4.70).

Before we can compare the interval and power-based approaches to sinusoid detection, it is necessary to formulate a new squared-envelope detector to use as a baseline. Specifically, the likelihood function $p_E(e \mid H_1)$ must model the squared-envelope of an additive mixture of a sinusoid and a Gaussian process, rather than a mixture of two Gaussian processes. The relevant distributions are widely published (originally, Rice, 1944), and the main results are summarised below without derivation.

Sinusoid Detection based on the Squared Envelope

We recall from Section 4.2 that the distribution of the squared-envelope of a wide-sense stationary Gaussian noise process is described by the exponential p.d.f.,

$$p_E(e;\sigma^2) = \frac{1}{2\sigma^2}\exp\left(\frac{e}{-2\sigma^2}\right), \tag{5.61}$$

where $\sigma^2 \equiv \gamma(0)$, and $\gamma(\cdot)$ is the process autocovariance function.

Following the addition of a sinusoid of amplitude A to the mixture, the probability density function governing the squared-envelope becomes

$$p_E(e;\sigma^2,A) = \frac{1}{2\sigma^2}\exp\left(\frac{e+|\mathscr{H}_a(\omega_c)|^2A^2}{-2\sigma^2}\right) I_0\left(\frac{|\mathscr{H}_a(\omega_c)|A\sqrt{e}}{\sigma^2}\right), \tag{5.62}$$

where ω_c is the signal frequency, $\mathscr{H}_a(\cdot)$ is the frequency response of the analysis filter and $I_0(\cdot)$ is the modified Bessel function with order zero[1]. This probability density function is associated with the *non-central chi-squared distribution*[2] and its derivation in this context is explained in Whalen (1971). In closing, we note that $I_0(z) \rightarrow 1$ as $z \rightarrow 0$, which reassures us that if either A or $\mathscr{H}_a(\omega_c)$ is very small, corresponding to the case of a weak signal or severe attenuation, respectively, the signal-and noise density (5.62) approaches the noise-only density (5.61).

5.2.7 Experimental Results and Analysis

Two interval-based sinusoid detectors have been constructed and evaluated, along with an optimal sinusoid detector that operates on a sample of the squared-envelope. The detection tasks are the same as those carried out in Chapter 4, except that the signal is now a sinusoid rather than a narrow notch of noise. The results are plotted in Figure 5.11. The test statistic supplied to both interval detectors is computed by differencing two linearly-interpolated zero crossing times (*cf.* Figure 4.16). The first detector models the distribution of the intervals of a sinusoid in noise by assuming that a sinusoid can be modelled as a Gaussian process with vanishingly narrow bandwidth, according to Section 5.2.2. The second detector models the distribution of the intervals using the Rayleigh-sum approach described in Section 5.2.5. Theoretical considerations and the probability density functions plotted in Figure 5.10 both suggest that the second interval detector will outperform the first.

Two comments are in order concerning the results presented in Figure 5.11. First, any significant benefit of using the Rayleigh-sum approach instead of the limiting approach appears to be restricted to signal-to-noise ratios in excess of 20 dB. At lower SNRs,

[1]The integral definition according to Abramowitz and Stegun (1972, 9.6.16) is

$$I_0(z) = \frac{1}{\pi}\int_0^{\pi} e^{\pm z\cos\theta}\,d\theta.$$

[2]or alternatively, a modified version of the *Rician distribution*.

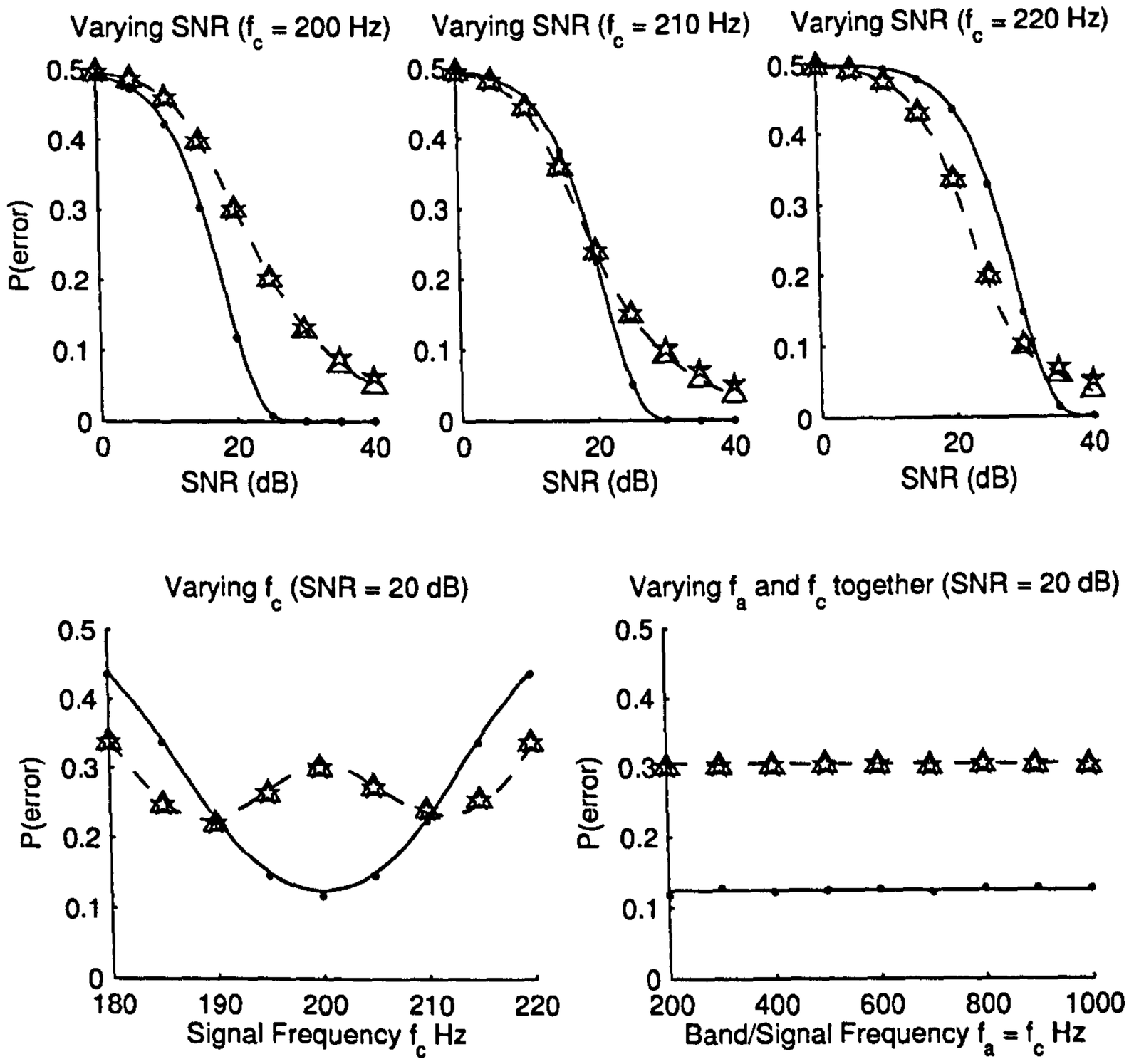


Figure 5.11: Probability of error for the interval-based sinusoid detectors. The predicted and observed values are shown using lines and markers, according to the following key: squared-envelope detector based on non-central chi-squared distribution (solid line; solid circle •); sinusoid interval detector using limits (no analytical results; pentagram) and sinusoid detector using sum of Rayleigh densities (dashed line; triangle △).

treating a sinusoid as a Gaussian process incurs very little penalty—in terms of the difference in the probability of error visible in the figure, at least—and could be an acceptable compromise. Second, and incidentally, comparing these results with those in Figure 4.18, it appears that the probability of error is significantly lower when the target signal is a sinusoid rather than a narrowband Gaussian process, despite the fact that the signal-to-noise ratio is identical. Most of this discrepancy arises because the results are given with respect to the pre-analysis SNR, which does not take into account the effect of the analysis filter on the signal, as opposed to the post-analysis SNR, which does. (The difference was explained in Section 4.1.3.)

5.3 Combining Power and Timing Detection

All the detectors described so far may be assigned to one of two categories: detection based exclusively on *power*, and detection based exclusively on *timing*. In an effort to contrast the two modes of detection, they have been discussed in separate sections and evaluated competitively; but this does not imply that power and timing detection are irreconcilable. In fact, we may recall from Chapter 2 that the duplex theory of auditory processing, which tentatively reserves a role for both average-rate and timing mechanisms, receives considerable support in the physiological literature and has consequently found expression in several computational models, including the EIH and ZCPA. Kim et al. (1999) have shown that incorporating intensity information into a signal representation based solely on zero crossings improves the performance of a speech recogniser. The aim of this section is to develop a detector that processes information from both the envelope and the zero crossing intervals—if possible, in an optimal fashion—as a step towards assessing the detectability of signals in joint representations such as the ZCPA.

The joint detectors described next combine two optimal detectors from the previous chapter. The first branch consists of the continuous interval detector, chosen because it outperforms the sampled interval detector and is more elegant than the interpolated interval detector. The second branch consists of a squared-envelope detector, chosen on account of its ubiquity and low-SNR performance (Whalen, 1984). The likelihood test for the minimum-error joint interval-peak detector uses the ratio of the joint conditional density functions:

$$\text{choose } H_1 \text{ iff } \frac{p_{I_cE}(i_c, e \mid H_1)}{p_{I_cE}(i_c, e \mid H_0)} > 1, \text{ otherwise choose } H_0. \tag{5.63}$$

In (5.63), I_c is the random variable governing the time between successive zero crossings of a continuous random process, and a continuous interval observation, denoted i_c, is computed by differencing the zero crossing times of a sampled process estimated via linear interpolation. E is the random variable governing the squared-envelope of the process. In what follows, we measure the peak squared-amplitude across an interval instead of the peak squared-envelope, which is harder to compute, but continue to denote the measurement using e. The justification for this approximation is explored more closely in Section 5.3.4.

5.3.1 Naïve Joint Interval-Peak Detector

The naïve approach to modelling a joint density funtion is to replace it with the product of the marginal densities, which assumes, often incorrectly, that the variables concerned are statistically independent. In the present case, we assume that the duration of a continuous interval and its peak squared-amplitude are independent, allowing us to write

$$p_{I_cE}(i_c, e \mid H_j) = p_{I_c}(i_c \mid H_j) p_E(e \mid H_j). \tag{5.64}$$

The marginal conditional p.d.f.s, $p_{I_c}(\cdot)$ and $p_E(\cdot)$, are readily available, so the naïve joint interval-peak detector can in principle be constructed without further effort.

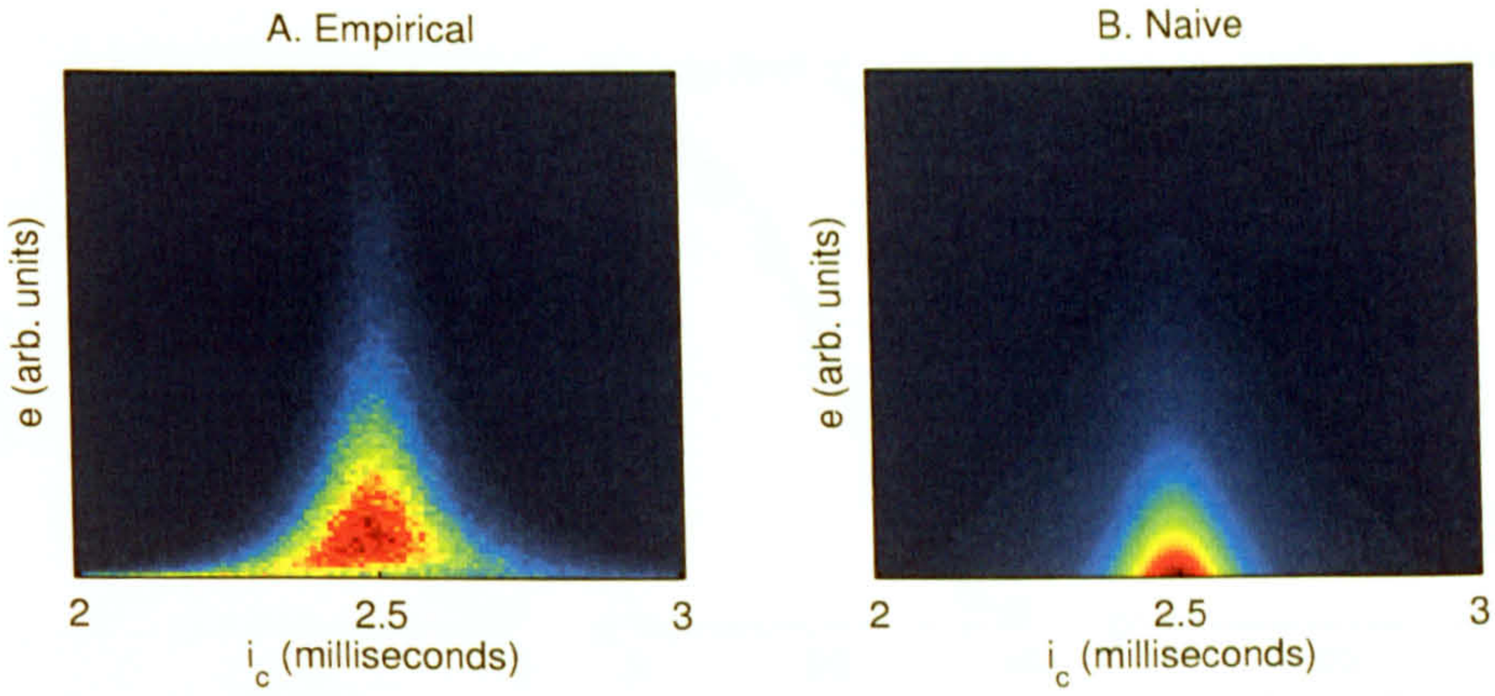


Figure 5.12: A) empirical joint probability density function governing interval duration and peak squared-amplitude, based on 250,000 samples and plotted as an image; B) the naïve joint p.d.f. formed by a product of marginals.

The empirical joint probability density function governing the envelope is shown in Figure 5.12A next to the naïve analytical version in Figure 5.12B. The two surfaces differ considerably, but there is some resemblance in the broad, triangular shape.

The performance of the naïve joint interval-peak detector has been evaluated under the same experimental conditions as the detectors presented in earlier sections, and the results are shown in Figure 5.13. The graphs show the change in error associated with i) changes in signal-to-noise ratio (top row of plots), ii) moving the signal away from the band centre whilst the SNR prior to analysis is fixed at -20 dB (bottom-left plot), and iii) changing both the band centre frequency and signal frequency together (bottom-right plot). Because the true joint distribution of the intervals and peaks is not yet known, it is impossible to predict the performance of the naïve joint detector, so only empirical results are shown. The joint detector outperforms the interval and envelope detectors under all conditions, except those in which the signal is centred on the band. For signals at the centre of the analysis band, the probability of error exceeds that associated with the envelope detector operating on its own, despite the fact that the test statistic supplied to the joint detector is augmented with information besides the peak squared-amplitude (namely, the interval duration). Along with the comparison between the probability density functions, this result confirms that the naïve assumption is inadequate and motivates the search for a better solution.

5.3.2 Capturing the Statistical Dependency between Intervals and Peaks

The failure of the naïve detector can be traced to the fact that a zero crossing interval and its peak squared-amplitude are statistically dependent. We can make some progress towards modelling this statistical dependency by assuming that the peak across an interval occurs exactly halfway between the zero crossings and then considering how the samples in a Gaussian process can be conditioned.

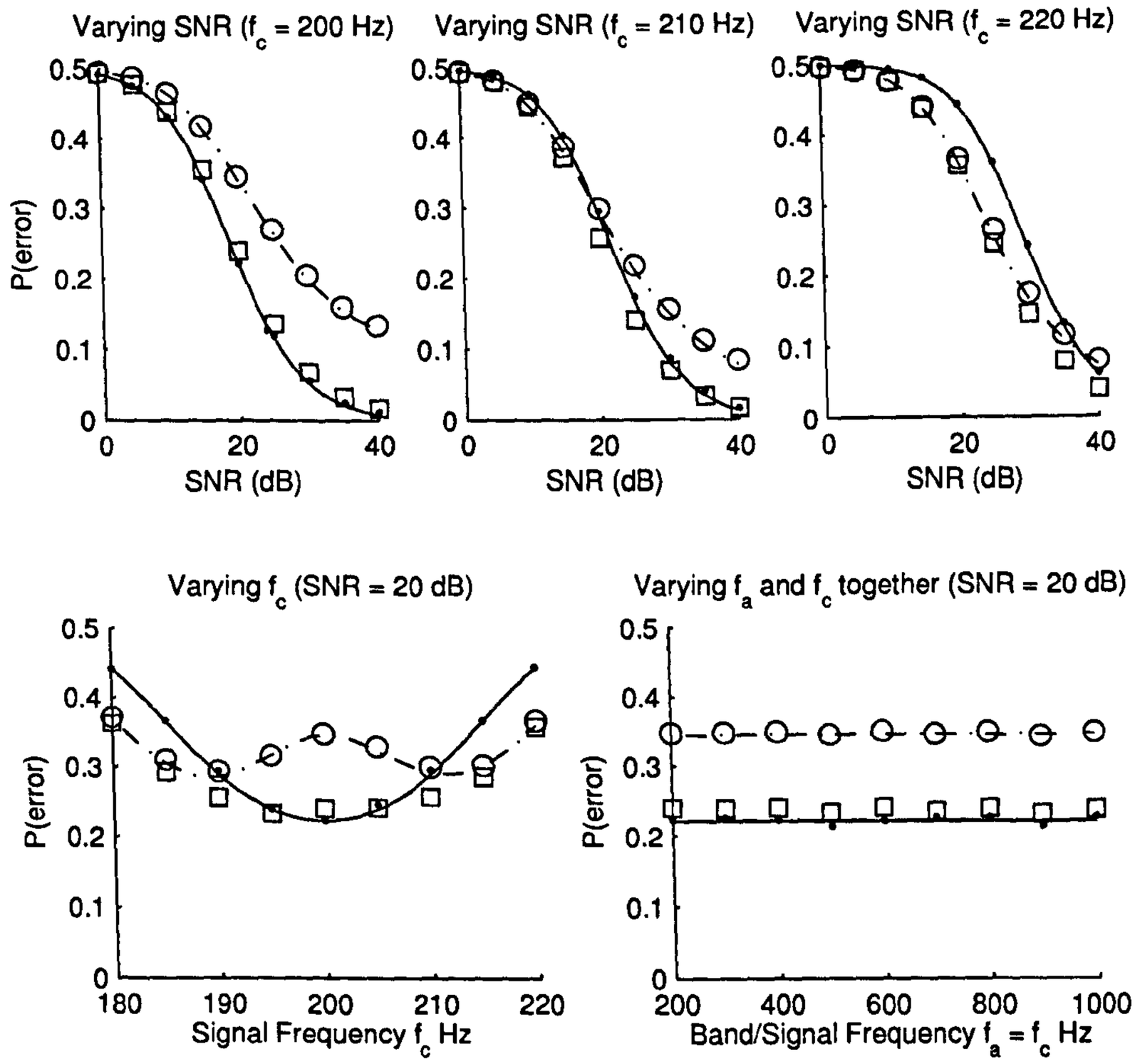


Figure 5.13: Probability of error in the naïve joint interval-peak detector. The predicted and observed values are shown using lines and markers, according to the following key: squared-envelope detector (solid line; solid circle •); continuous interval detector (dash-dotted; open circle ○) and naïve joint interval-peak detector (no analytical results; open square □).

Conditioning on Zero Values

Our key aim in this section is to find the probability density function governing the square of a sample in a Gaussian process, X, given that it is observed in the middle of a zero crossing interval of duration i_c. Three samples of a zero-mean stationary Gaussian process, when separated by intervals of $\frac{1}{2}i_c$ (see Figure 5.14A) are governed by the density function

$$p_{\mathcal{A}\mathcal{X}}(A, x_1, x_2; i_c) = \frac{1}{(2\pi)^{3/2}|\Sigma_1|^{1/2}} \exp \frac{\mathbf{x}^T \Sigma_1^{-1} \mathbf{x}}{-2}, \tag{5.65}$$

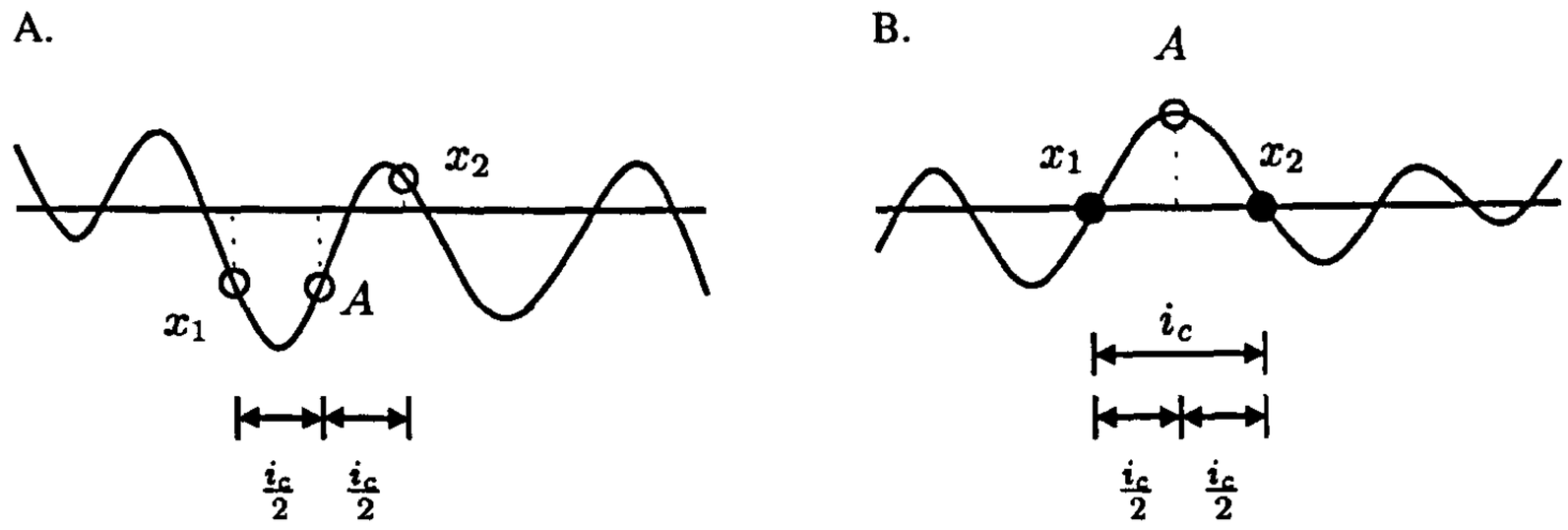


Figure 5.14: A) Three samples are separated by $i_c/2$, and candidate waveforms are unconstrained. B) When the first and third samples are zero, candidate waveforms contain a zero crossing interval of duration i_c and, in narrowband processes, the middle sample A is a nominal peak (or trough).

where $\mathbf{x} \triangleq (A, x_1, x_2)^T$ and, with γ_X denoting the autocovariance function,

$$\Sigma_1 = \begin{pmatrix} \gamma_X(0) & \gamma_X(\frac{1}{2}i_c) & \gamma_X(\frac{1}{2}i_c) \\ \gamma_X(\frac{1}{2}i_c) & \gamma_X(0) & \gamma_X(i_c) \\ \gamma_X(\frac{1}{2}i_c) & \gamma_X(i_c) & \gamma_X(0) \end{pmatrix}. \tag{5.66}$$

The p.d.f. in (5.65) describes the distribution that one would expect three samples of a Gaussian process to follow in the absence of any constraint besides the separation in time prescribed by i_c.

In the next step, we modify the probability density function by securing the first and last sample to the time axis by conditioning on $X_1 = X_2 = 0$, as shown in Figure 5.14B. The p.d.f.[1] that results is a function of the middle sample alone, or nominal peak amplitude, namely,

$$\tilde{p}_{A|I_c}(A \mid i_c) = \frac{1}{(2\pi|\Sigma_2|)^{1/2}} \exp \frac{A^2}{-2\Sigma_2}, \tag{5.67}$$

in which

$$\Sigma_2 = \frac{\gamma_X^2(0) + \gamma_X(0)\gamma_X(i_c) - 2\gamma_X^2(\frac{1}{2}i_c)}{\gamma_X(0) + \gamma_X(i_c)}. \tag{5.68}$$

Hence we have shown that the question, "If two samples in a Gaussian process are observed to be zero, what distribution governs the sample halfway between them?" has the answer, "A Gaussian distribution with zero mean and variance Σ_2." One final step remains: A refers to the value of a Gaussian sample, whereas our goal is to find the distribution governing its *square*. Using the replacement $E = A^2$, the new conditional p.d.f. is found to be

$$\tilde{p}_{E|I_c}(e \mid i_c) = \frac{1}{(2\pi e|\Sigma_2|)^{1/2}} \exp \frac{e}{-2\Sigma_2}. \tag{5.69}$$

[1] The tilde above $\tilde{p}$ is a reminder that the p.d.f. has been approximated.

Representing Zero Crossings by Naïvely Conditioning on Zeros

We have now obtained an expression for the probability density function governing the (squared) sample at the midpoint between two zero values in a Gaussian random process. From this result, one might be led to reason as follows.

1. From the previous chapter, we understand that the probability density associated with the observation of an interval of duration i_c, i.e., $p_{I_c}(i_c)$, can be computed.

2. The observation of an interval of duration i_c implies two zeros in the signal at times $t_0 - \frac{1}{2}i_c$ and $t_0 + \frac{1}{2}i_c$, with sample values denoted x_1 and x_2, respectively. The midpoint of the interval falls at time t_0, and we label its value A.

3. Because x_1 and x_2 are fixed, we can determine the probability density function that describes the square of the midpoint, i.e., E, by application of (5.69). The joint p.d.f. is then given by the product of the marginal and conditional densities:

$$p_{I_c E}(i_c, e) \approx p_{I_c}(i_c)\tilde{p}_{E|I_c}(e \mid i_c). \tag{5.70}$$

This route to finding the joint probability density function is invalid for reasons that will be addressed after a couple of remarks in connection with the immediate result. First, this expression does fulfil the two basic requirements of a probability density function, namely, it is nonnegative for all pairs (i_c, e) and the total volume under the p.d.f. is one. Second, and by definition, one obtains the correct marginal p.d.f. for i_c by integration, i.e.,

$$p_{I_c}(i_c) = \int_0^\infty p_{I_c E}(i_c, e)de,$$

though the same does not hold for the marginal p.d.f. in E. The probability density function $p_{I_c E}(\cdot)$ is shown in Figure 5.15B, alongside the empirical version. The two images are clearly quite dissimilar, which indicates a fault in the latest approach.

Adjusting for Differential Areas when Representing Zero Crossings

The failure of the zero-conditioning approach presented above stems from the conflation of two quantities: the probability density associated with X_1 and X_2, and that associated with I_c. Properly speaking, the function $p_{\ldots\mathcal{X}}(\ldots, x_1{=}0, x_2{=}0)$ expresses the probability that the sample values x_1 and x_2 lie in a vanishingly small region close to zero (in conjunction with whatever replaces the ellipses); the difference between the sample times is fixed at i_c. The function $p_{I_c}(i_c)$ bears the converse description: it expresses the probability that the difference in sample times occupies a vanishingly small region surrounding i_c, whilst the values of the samples are fixed at zero. Considering one of the crossings, a small region of probability density associated with a small change in X around zero, which we label δX, corresponds to the probability density associated with a small change in interval duration, δI, as shown in Figures 5.16A and 5.16B. It is evident from the diagram, and the discussion of perturbation analysis in Section 1.2.2, that the ratio between the two depends on the slope of the signal through the zero crossing.

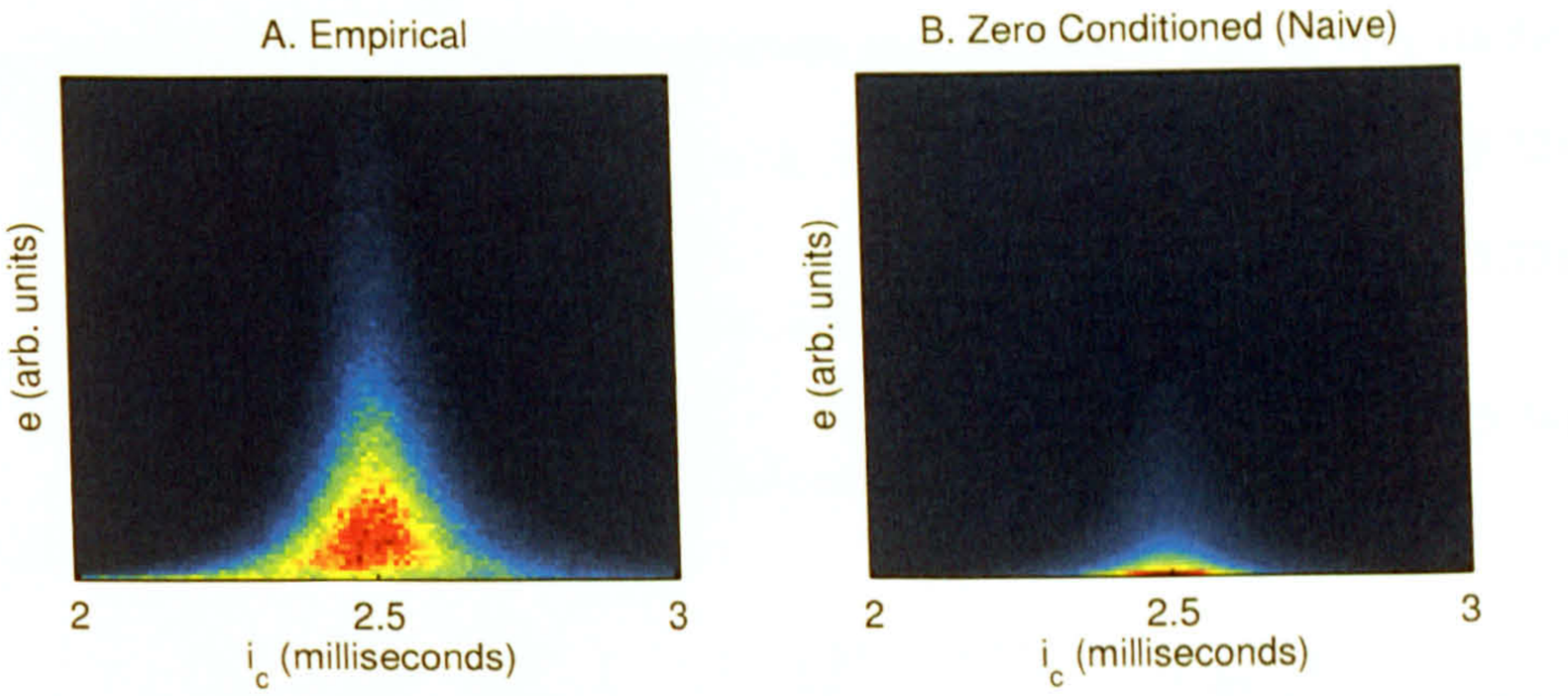


Figure 5.15: A) empirical joint probability density function governing interval duration and peak squared-amplitude, based on 250,000 samples and plotted as an image; B) the joint p.d.f. formed by conditioning the peak sample on zero values either side (eqn. 5.70).

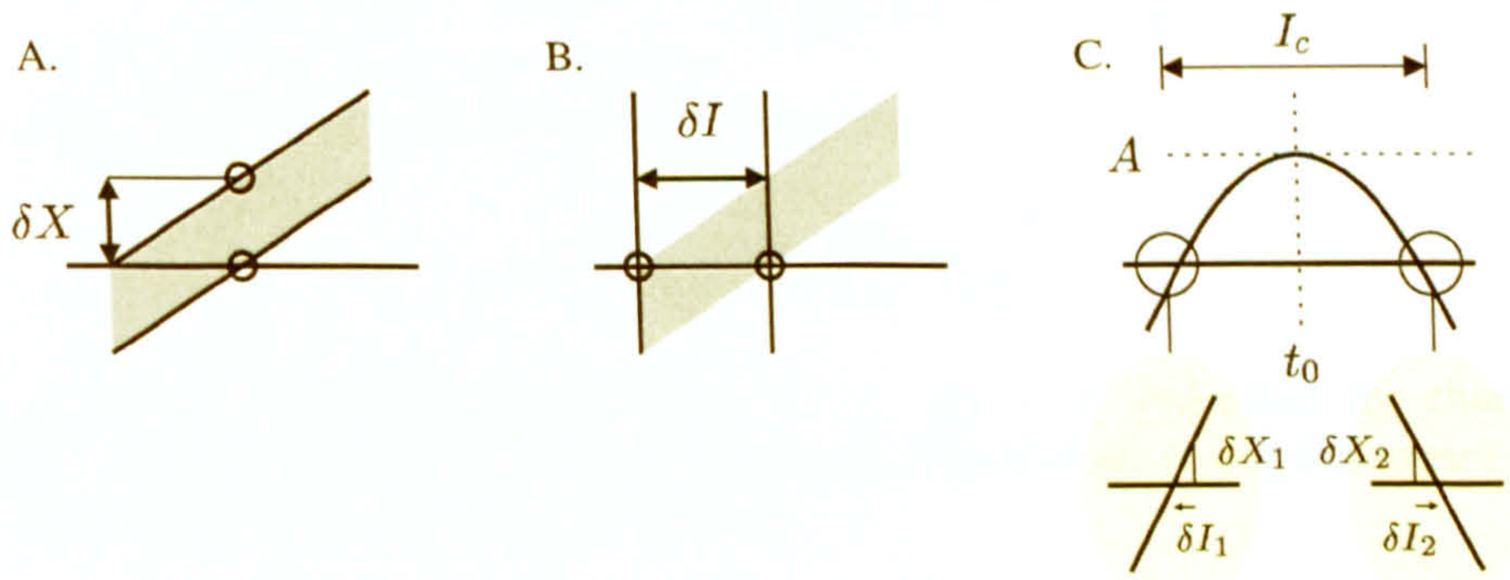


Figure 5.16: A) a small variation in a sample value around zero; B) the corresponding small variation in the zero crossing time, dependent upon the slope of the waveform near the crossing; C) an estimate of the slope of the waveform is found by stereotyping the waveform to a cosine period around t_0.

The derivatives at the zero crossing times must be known before the probability density functions can be corrected; unfortunately, they are not. This problem is addressed by stereotyping the waveform on very short timescales to a sinusoidal form. (See Figure 5.16C.) This seems to be a reasonable assumption for a narrowband process with a slow-varying envelope. If the nominal (signed) peak value, A, occurs at time t_0, and is preceded and followed by zero crossings at times $t_0 - \frac{1}{2}i_c$ and $t_0 + \frac{1}{2}i_c$, respectively, then the signal is approximated locally by

$$x(t) \approx A \cos\left(\frac{\pi(t - t_0)}{i_c}\right). \tag{5.71}$$

Using this relationship, it is possible to eliminate any reference to δX_1 or δX_2 via the following change of variables

$$A = A \tag{5.72}$$

$$\delta X_1 = \delta I_1 \left.\frac{dx}{dt}\right|_{t=t_0-\frac{1}{2}i_c} = \delta I_1 \frac{A\pi}{i_c} \tag{5.73}$$

$$\delta X_2 = -\delta I_2 \left.\frac{dx}{dt}\right|_{t=t_0+\frac{1}{2}i_c} = \delta I_2 \frac{A\pi}{i_c} \tag{5.74}$$

which has the Jacobian determinant

$$J = \begin{vmatrix} \frac{\partial A}{\partial A} & \frac{\partial \delta X_1}{\partial A} & \frac{\partial \delta X_2}{\partial A} \\ \frac{\partial A}{\partial \delta I_1} & \frac{\partial \delta X_1}{\partial \delta I_1} & \frac{\partial \delta X_2}{\partial \delta I_1} \\ \frac{\partial A}{\partial \delta I_2} & \frac{\partial \delta X_2}{\partial \delta I_2} & \frac{\partial \delta X_2}{\partial \delta I_2} \end{vmatrix} = \begin{vmatrix} 1 & \frac{\pi}{i_c}\delta I_1 & \frac{\pi}{i_c}\delta I_2 \\ 0 & \frac{A\pi}{i_c} & 0 \\ 0 & 0 & \frac{A\pi}{i_c} \end{vmatrix} = \left(\frac{A\pi}{i_c}\right)^2. \tag{5.75}$$

The transformed probability density function is

$$p_{\mathcal{AI}}(A, \delta i_1, \delta i_2; i_c) = \frac{A^2\sqrt{2\pi}}{4i_c^2|\Sigma_1|^{1/2}} \exp \frac{\mathbf{y}^T\Sigma_1^{-1}\mathbf{y}}{-2}. \tag{5.76}$$

where Σ_1 is defined in the same way as (5.66), and

$$\mathbf{y} \triangleq \left(A, \frac{A\pi}{i_c}\delta i_1, \frac{A\pi}{i_c}\delta i_2\right)^T.$$

We can now condition appropriately on $\delta I_1 = \delta I_2 = 0$, and effect the change of variables $E = A^2$, so that the probability density function governing the square of the nominal peak is given by the approximation

$$\tilde{p}_{E|I_c}(e \mid i_c) = \sqrt{\frac{e}{2\pi|\Sigma_2|^3}} \exp \frac{e}{-2\Sigma_2}, \tag{5.77}$$

where the definition of Σ_2 has not changed from (5.68). The adjusted joint p.d.f. is once again given by the product of the conditional and marginal density functions, this time using the expression for the conditional p.d.f. derived in (5.77):

$$p_{I_cE}(i_c, e) \approx p_{I_c}(i_c)\tilde{p}_{E|I_c}(e \mid i_c). \tag{5.78}$$

Figures 5.17A and 5.17B provide an image of the empirical probability density function and the analytical p.d.f. according to $p_{I_cE}(\cdot)$, respectively. The visible similarity between the images—including the slight asymmetry—is encouraging, and suggests that an optimal detector that incorporates these likelihood functions will outperform the naïve version, and possibly both marginal detectors.

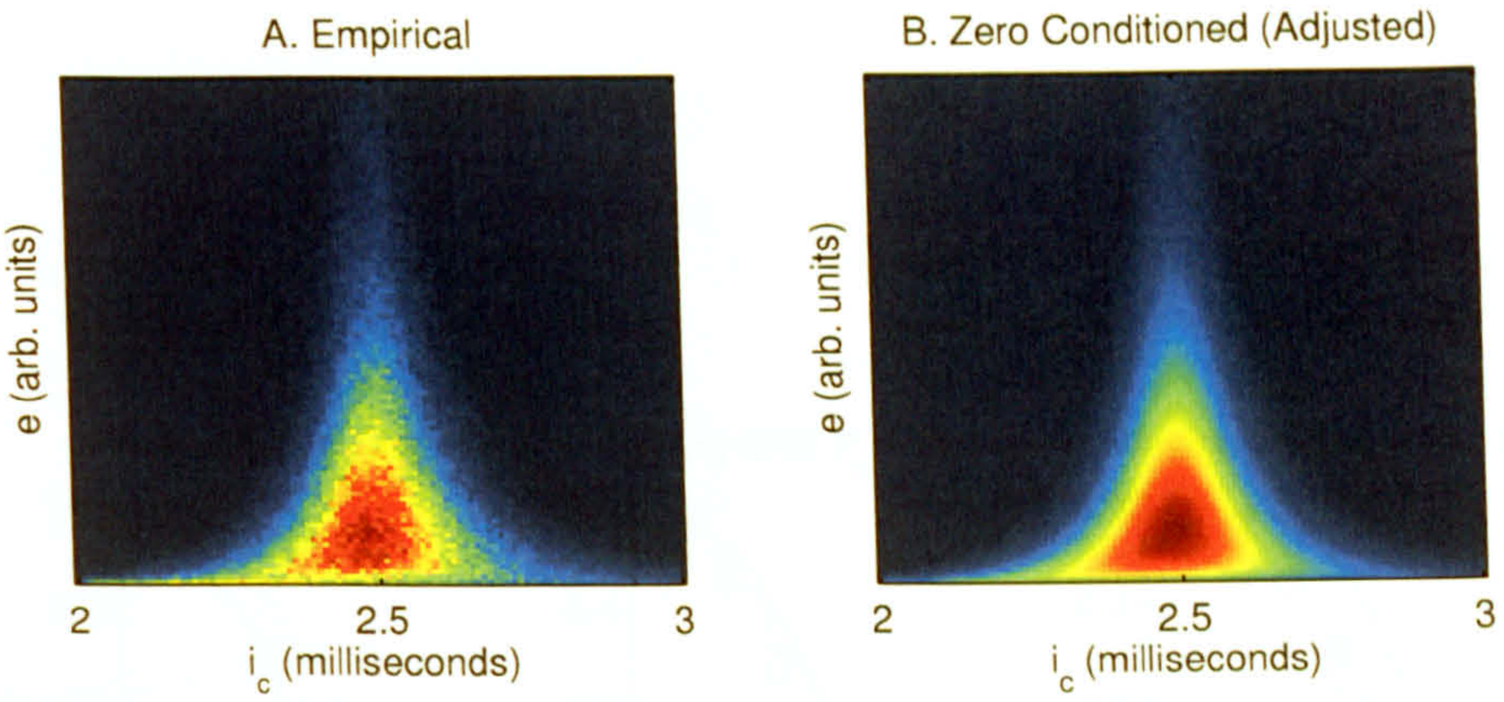


Figure 5.17: A) empirical joint probability density function governing interval duration and peak squared-amplitude, based on 250,000 samples and plotted as an image; B) the analytical joint p.d.f., which takes into account the rescaling of differential areas (eqn. 5.78).

5.3.3 Experimental Results and Analysis

The adjusted joint interval-peak detector was constructed and evaluated according to a minimum error criterion. The likelihood functions are given by (5.78) and placed into the decision rule (5.63). The empirical results are plotted in Figure 5.18, along with the results for the naïve interval detector, and the squared-envelope and continuous interval detectors. The adjusted joint interval-peak detector outperforms the other detectors when the signal is displaced from the band centre. When the signal is centred on the band, the performance of the joint detector matches that of the squared-envelope detector.

It is notable that the error curve for the adjusted joint interval-peak detector falls below the convex hull formed by the error curves of the individual continuous interval and squared-envelope detectors. (The same can be said for the naïve joint detector at some data points.) Hence, the duration of a zero crossing interval provides information that the peak amplitude across that interval does not, and *vice versa*. This result is quite notable in that it shows that the information contributed by i_c still reduces the probability of error significantly, even when the signal is located near (but not precisely on) the band centre.

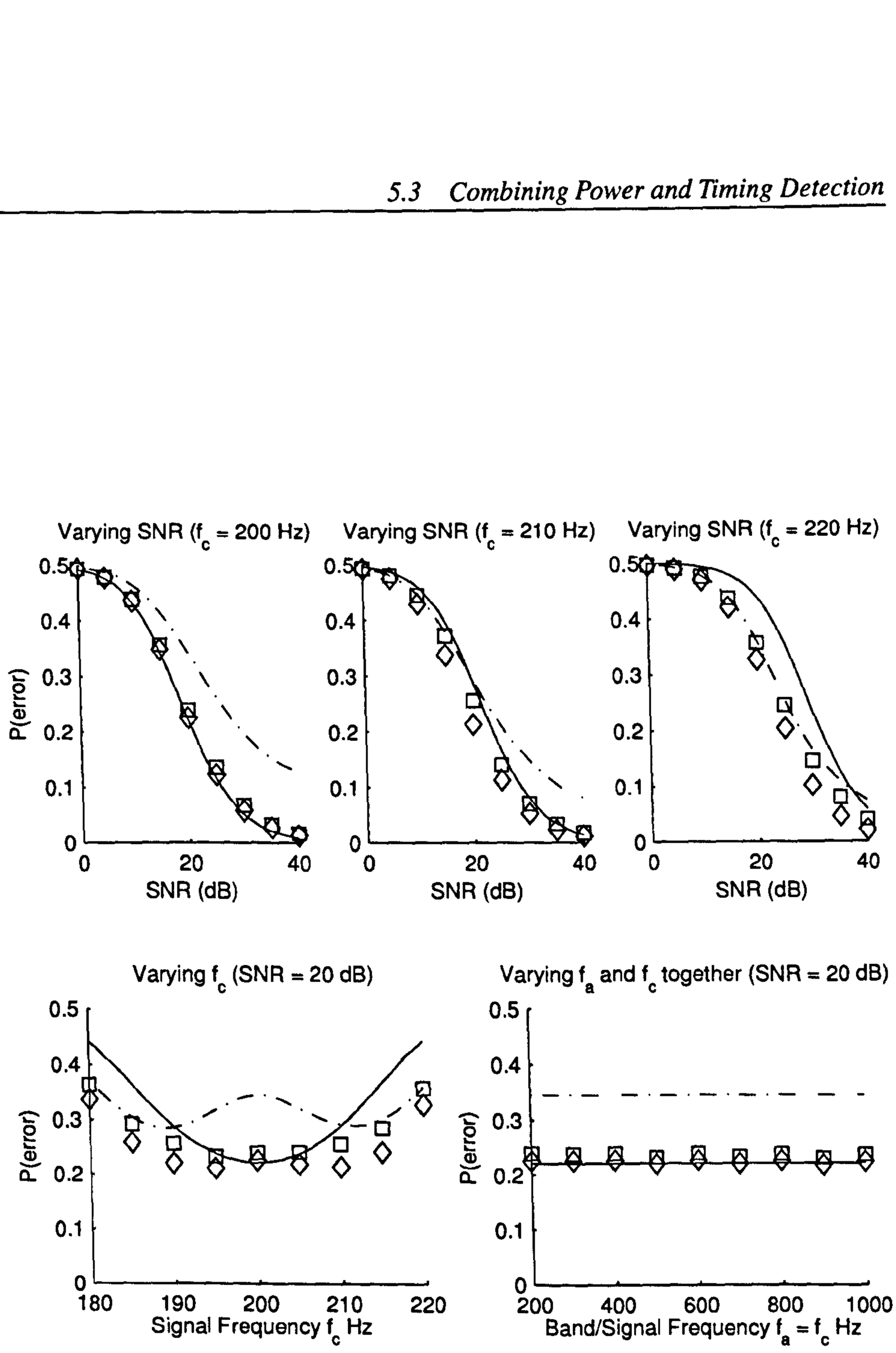


Figure 5.18: Probability of error in the adjusted joint interval-peak detector. The predicted and observed values are shown using lines and markers, according to the following key: squared-envelope detector (solid line; no marker); continuous interval detector (dash-dotted line; no marker), naïve joint interval-peak detector (no analytical results; open square □); and adjusted joint interval-peak detector (no analytical results; diamond ◇).

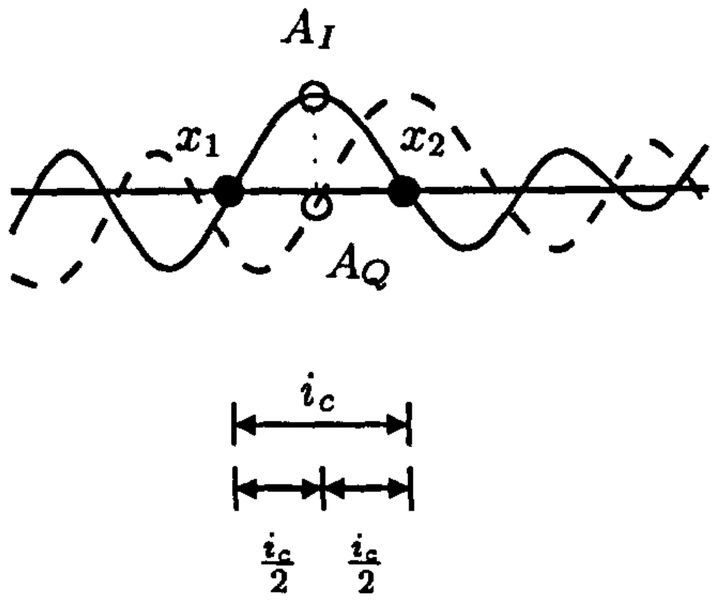


Figure 5.19: Four samples in an analytical time-domain signal. The samples x_1, x_2 and A_I appear in the in-phase component (solid line); the fourth, A_Q, appears in the quadrature component (dashed line). The samples A_I and A_Q coincide at time t and together determine the instantaneous envelope and phase of the nominal peak. The samples x_1 and x_2 are fixed at zero. (See also Figure 5.14.)

5.3.4 Conditioning the Squared-Envelope on Zero Crossings

In order to simplify the working, we have assumed throughout this section that the squared-envelope can be modelled by the square of the signal amplitude halfway between the zero crossings. In these closing remarks, we shall derive results concerning the distribution of the actual squared-envelope, in order to determine whether the squared-amplitude is a suitable replacement. (We still assume that the peak falls exactly halfway between the zero crossings.)

The squared-envelope, $e(t)$, of an analytical signal is defined as

$$e(t) = x_I^2(t) + x_Q^2(t) \tag{5.79}$$

where x_I and x_Q are its in-phase and quadrature components. It is governed by the random variable E. We shall designate the in-phase sample and the quadrature sample at time t, A_I and A_Q, respectively. Let us in addition identify two samples in the in-phase signal, x_1 and x_2, which occur at times $t - \frac{1}{2}i_c$ and $t + \frac{1}{2}i_c$, respectively. The four samples $\mathbf{x} \triangleq (A_I, A_Q, x_1, x_2)^T$ are joint Gaussian-distributed with covariance matrix[1]

$$\Sigma_3 = \begin{pmatrix} \gamma(0) & 0 & \gamma(\frac{1}{2}i_c) & \gamma(\frac{1}{2}i_c) \\ 0 & \gamma(0) & \hat{\gamma}(-\frac{1}{2}i_c) & \gamma(\frac{1}{2}i_c) \\ \gamma(\frac{1}{2}i_c) & \hat{\gamma}(-\frac{1}{2}i_c) & \gamma(0) & \gamma(i_c) \\ \gamma(\frac{1}{2}i_c) & \hat{\gamma}(\frac{1}{2}i_c) & \gamma(i_c) & \gamma(0) \end{pmatrix}. \tag{5.80}$$

Notice that we are pursuing a similar strategy to before—conditioning the midpoint on the presence of zero values either side—only this time we are taking into account the

[1] The subscripts on γ have been dropped.

contribution of the quadrature component to the envelope. The relevant quantities are sketched in Figure 5.19.

The autocovariance function $\gamma(\tau)$ is defined as usual, with the exception that we have made explicit the fact we are referring to the covariance of two samples of the in-phase signal:

$$\gamma(\tau) = E\{x_I(t)x_I(t-\tau)\}.$$

We have also defined a modified autocovariance function, $\hat{\gamma}(\cdot)$, which the computes the covariance of two samples, if the first appears in the in-phase signal, and the second in the quadrature signal:

$$\hat{\gamma}(\tau) = E\{x_I(t)x_Q(t-\tau)\}.$$

Note that $\hat{\gamma}(0) = 0$, and $\hat{\gamma}(\tau) = -\hat{\gamma}(-\tau)$.

It can be shown that, when A_I and A_Q are conditioned on $X_1 = X_2 = 0$, they are joint Gaussian-distributed, with zero mean and diagonal covariance matrix,

$$\Sigma_4 = \begin{pmatrix} \gamma(0) - \dfrac{2\gamma^2(\frac{1}{2}i_c)}{\gamma(0)+\gamma(i_c)} & 0 \\ 0 & \gamma(0) - \dfrac{2\hat{\gamma}^2(\frac{1}{2}i_c)}{\gamma(0)-\gamma(i_c)} \end{pmatrix} \equiv \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}^{-1}. \quad (5.81)$$

If the diagonal elements of the inverse matrix, Σ_4^{-1}, are individually labelled α and β, then the joint density function of A_I and A_Q can be written

$$p_{A_I A_Q}(A_I, A_Q) = \frac{|\alpha\beta|^{1/2}}{2\pi} \exp\left(\frac{\alpha A_I^2 + \beta A_Q^2}{-2}\right). \quad (5.82)$$

Performing the change of variables

$$A_I = \sqrt{E}\cos\theta \quad (5.83)$$
$$A_Q = \sqrt{E}\sin\theta \quad (5.84)$$

in (5.82) provides a new distribution in terms of the squared-envelope, E, and instantaneous phase, Θ:

$$p_{E\Theta}(e,\theta) = \frac{|\alpha\beta|^{1/2}}{4\pi} \exp\left[\frac{e(\alpha\cos^2\theta + \beta\sin^2\theta)}{-2}\right]. \quad (5.85)$$

Lastly, we integrate to obtain the marginal probability density function that governs the squared-envelope, irrespective of phase.

$$\begin{aligned} p_E(e) &= \frac{|\alpha\beta|^{1/2}}{4\pi} \exp\left[\frac{e(\alpha+\beta)}{-4}\right] \int_0^{2\pi} \exp\left[\frac{e(\beta-\alpha)\cos 2\theta)}{4}\right] d\theta \\ &= \frac{|\alpha\beta|^{1/2}}{2} \exp\left[\frac{e(\alpha+\beta)}{-4}\right] I_0\left[\frac{e(\beta-\alpha)}{4}\right]. \end{aligned} \quad (5.86)$$

In summary, (5.86) is the probability density function governing the squared-envelope of the midpoint between two sample values, if those values occupy a vanishingly small region surrounding zero.

No conditioning

Consider the case in which there is no zero value conditioning, or the conditioned samples are independent of the mid-point sample. In this case, it is easy to verify that

$$\alpha = \beta = \frac{1}{\gamma(0)} \equiv \frac{1}{\sigma^2},$$

and (5.86) reduces to

$$p_E(e) = \frac{1}{2\sigma^2} \exp\left(\frac{e}{-2\sigma^2}\right). \tag{5.87}$$

Thus our expression for the squared-envelope correctly reduces to the exponential probability density function that governs an individual sample of the squared-envelope of a Gaussian process.

Conditioning with a Narrowband Assumption

Second, consider the narrowband case in which the interval duration, i_c, is close to the zero crossing interval of the channel centre frequency, i.e., $2i_c \approx 1/f_a$. Under these circumstances, $\gamma(\frac{1}{2}i_c) \approx 0$ and $\hat{\gamma}(\frac{1}{2}i_c) \approx \gamma(0)$, and it may be shown that $1/\alpha = \Sigma_2$ and $1/\beta \approx 0$.

Because β is very large, the modified Bessel function, $I_0(\cdot)$ can be approximated by its first-order asymptotic series expansion[1] (Abramowitz and Stegun, 1972, 9.7.1),

$$I_0(z) \approx \frac{\exp(z)}{\sqrt{2\pi z}}. \tag{5.88}$$

Replacing $I_0(\cdot)$ with (5.88) in (5.86) gives

$$\tilde{p}_{E|I_c}(e \mid i_c) \approx \frac{|\alpha\beta|^{1/2}}{\sqrt{2\pi e(\beta-\alpha)}} \exp\left[\frac{e(\alpha+\beta)}{-4}\right] \exp\frac{e(\beta-\alpha)}{4} \tag{5.89}$$

$$= \left[2\pi e\left(\frac{1}{\alpha} - \frac{1}{\beta}\right)\right]^{-1/2} \exp\left(\frac{\alpha e}{-2}\right) \tag{5.90}$$

$$\approx \frac{1}{(2\pi e\Sigma_2)^{1/2}} \exp\left(\frac{e}{-2\Sigma_2}\right). \tag{5.91}$$

This conforms to the probability density function in (5.67).

Adjusting for Differential Areas

Thirdly, we consider the case in which the differential areas have been adjusted to account for the slope of the line through the zero crossings, having assumed that the

[1] For a discussion of asymptotic expansions consult Self (2005).

in-phase signal is locally sinusoidal and has a peak A_I at time t_0. In this case, it can be shown that

$$\tilde{p}_{E|I_c}(e \mid i_c) = \frac{e\sqrt{\alpha^3\beta}}{4} \exp\left[\frac{e(\beta+\alpha)}{-4}\right] \left\{ I_0\left[\frac{e(\beta-\alpha)}{4}\right] + I_1\left[\frac{e(\beta-\alpha)}{4}\right] \right\}, \tag{5.92}$$

where $I_1(\cdot)$ is the modified Bessel function of order one[1].

The asymptotic expansion of $I_1(\cdot)$ is the same as that of $I_0(\cdot)$, i.e.,

$$I_1(z) \approx \frac{\exp(z)}{\sqrt{2\pi z}} \tag{5.93}$$

(Abramowitz and Stegun, 1972, 9.7.1). Using (5.88) and (5.93) in (5.92) results in

$$\tilde{p}_{E|I_c}(e \mid i_c) \approx \alpha\sqrt{e}\left[2\pi\left(\frac{1}{\alpha}-\frac{1}{\beta}\right)\right]^{-1/2} \exp\left[\frac{\alpha e}{-2}\right]. \tag{5.94}$$

If we assume, as in the section above, that the interval durations are close to those of the channel centre frequency, i.e., $2i_c \approx 1/f_a$, such that $1/\beta \approx 0$, then

$$\tilde{p}_{E|I_c}(e \mid i_c) \approx \sqrt{\frac{e}{2\pi\Sigma_2^3}} \exp\left[\frac{e}{-2\Sigma_2}\right]. \tag{5.95}$$

This conforms to the probability density function in (5.77).

[1]The integral definition according to Abramowitz and Stegun (1972, 9.6.19) is

$$I_n(z) = \frac{1}{\pi}\int_0^{\pi} e^{z\cos\theta}\cos(n\theta)d\theta.$$

5.4 Detection using Multiple Intervals

The elementary interval detectors of the previous chapter focussed on detection using a single zero crossing interval received in a narrowband channel. This section addresses the detection of a signal using a number of consecutive intervals. Drawing again on the analogy between zero crossings and auditory nerve spikes, the multiple interval detector may be said to model a cell which responds to certain patterns of intervals.

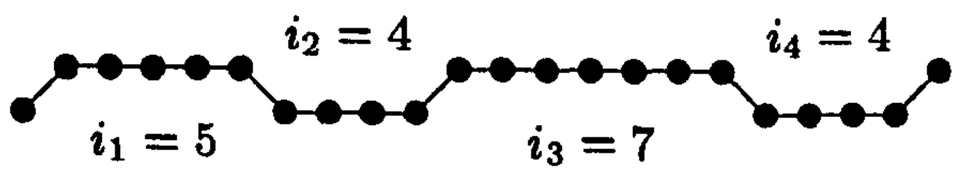
A sampled detector operating on N intervals employs the rule:

$$\text{choose } H_1 \text{ iff } \frac{p_{I_1 \ldots I_N}[i_1, i_2, \ldots, i_N \mid H_1]}{p_{I_1 \ldots I_N}[i_1, i_2, \ldots, i_N \mid H_0]} > 1, \text{ otherwise choose } H_0.$$

We have assumed here, without loss of generality, that the decision rule of the multiple interval detector minimises error, that is, a detection is made whenever the sequence of intervals is more likely given that the signal is present (H_1) than that it is absent (H_0). As usual, the first step is to determine the conditional probability density functions by which the detector makes decisions; in this case, we are interested in the joint p.d.f. governing N consecutive zero crossing intervals of a Gaussian process.

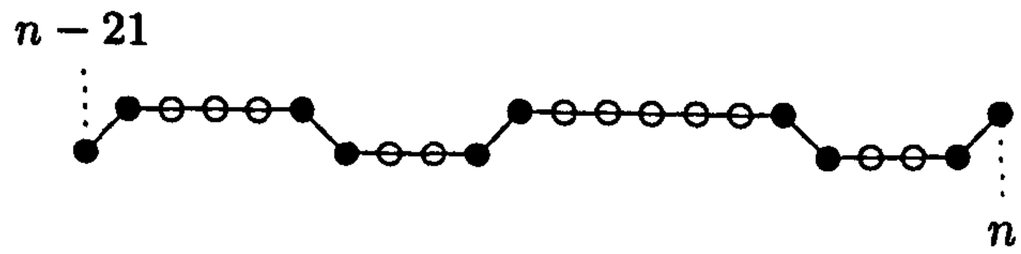
The multiple interval event is associated with a pattern of sign changes which coincide with the zero crossings. A multiple interval event consisting of N intervals is composed of $N + 1$ consecutive zero crossings with alternating direction. Provided that each interval is longer than one sample, we must determine the probability that $2N + 2$ samples are of a certain sign. This is the *dimensionality* of the problem, $d = 2N + 2$.

We can illustrate our technique with an example. Consider the binary time series sketched below.



This corresponds to the multiple interval event $\langle i_1 = 5, i_2 = 4, i_3 = 7, i_4 = 4 \rangle$. The probability of this interval sequence is equal to *twice* the probability of the pattern of sample signs shown above, as there is exactly one, equally-probable pattern of sign changes associated with this multiple interval event: namely, the same pattern with the zero crossings in the opposite direction.

If the octave constraint described in Chapter 4 is enforced, then we may marginalise to remove the samples which do not form part of a crossing, and consider only those that do.



If the process that generates the samples is Gaussian, then to find the probability of the multiple interval event $\langle i_1 = 5, i_2 = 4, i_3 = 7, i_4 = 4\rangle$, we only integrate over the samples $n, n{-}1, n{-}4, n{-}5, n{-}11, n{-}12, n{-}15, n{-}16, n{-}20$ and $n{-}21$; the signs of the other samples are implicit from the crossings. In this example, there are $N = 4$ intervals, and $d = 10$; thus the probability of this event is twice the ten-dimensional orthant probability

$$\frac{1}{(2\pi)^5|\Sigma|^{1/2}} \int_{-\infty}^{0} \cdots \int_{-\infty}^{0} \int_{0}^{\infty} \exp\left(\frac{\mathbf{x}^T\Sigma^{-1}\mathbf{x}}{-2}\right) dx_n dx_{n-1} \cdots dx_{n-21}, \quad (5.96)$$

where Σ is the 10×10 autocovariance matrix for the samples in the crossings. Pursuing this type of analysis—interpreting intervals as sign changes in a random process—will inevitably require the evaluation of d-dimensional orthant probabilities. Any orthant probability can be re-expressed in terms of a sum of all-positive orthant probabilities. Accordingly, in the sections that follow, we concentrate our effort on the more general problem of the all-positive orthant probability.

5.4.1 A Recursive Solution

Our first attempt to analyse the problem of the d-dimensional orthant probability examines the possibility of decomposing high-dimensional orthant probabilities into more manageable orthant probabilities of lower dimension, for which a solution is readily available, starting with the three dimensional case. Kedem (1980) explains the bridge between the two- and three-dimensional orthants algebraically (applying Boole's formula—see below). We shall now demonstrate the procedure using Venn diagrams.

The Three-dimensional Case

Consider three distinct samples, and let D_1, D_2 and D_3 respectively denote the event that the sample is non-negative. The Venn diagram in Figure 5.20A plots the three events as intersecting circles, dividing the total area into eight regions, each of which corresponds to a combination of events. The area of a region (schematically) relates the probability of its associated event, so the total area of the Venn diagram is one, i.e., a certain event. Thus, our interest—the all-positive orthant probability—is equal to the area of the central region.

In the remaining four Venn diagrams, we adopt the convention of labelling each region to keep track of how many times it has been counted; in Figure 5.20B, each region is counted once. With the knowledge that the total area is unity, our goal is to successively remove regions until only the central region remains. We first subtract the outer region, leaving only the region formed by the union of the three circles (Figure 5.20C), whose area is

$$1 - P(\bar{D}_1\bar{D}_2\bar{D}_3).$$

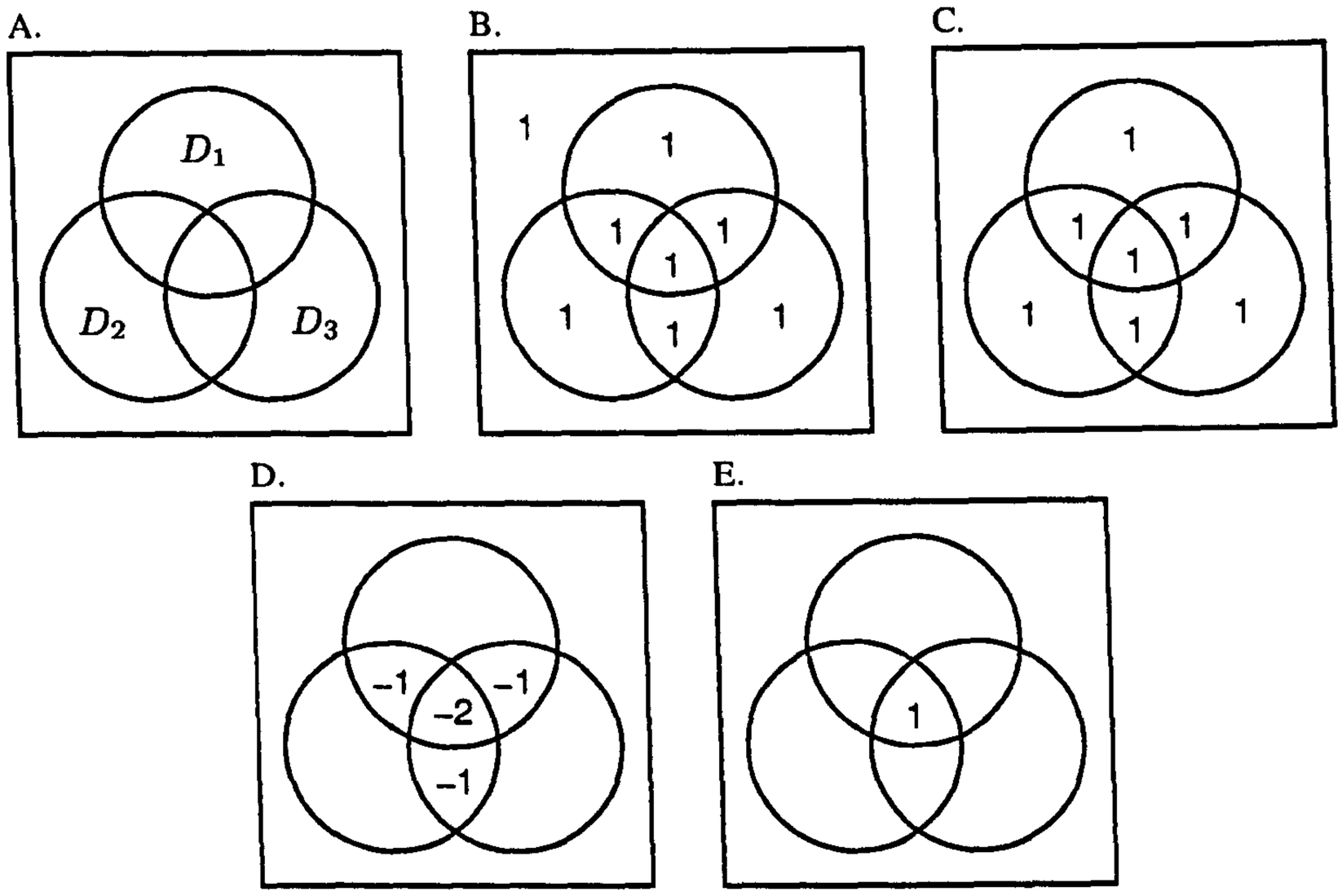


Figure 5.20: Relating 2D and 3D orthant probabilities via Venn diagrams. A) three events D_1, D_2 and D_3, drawn as intersecting circles; B) initially, every event is counted once; C) subtract the event that none of D_1, D_2 or D_3, occur; D) subtract each circle once, causing regions where two circles overlap to be subtracted twice, and the central region, where all three circles intersect, to be subtracted three times; D) add the regions where any two circles intersect back on, and only the central region remains.

An overbar denotes the complement of a region.[1]

Next, we subtract the region occupied by each circle, giving the area

$$1 - P(\bar{D}_1\bar{D}_2\bar{D}_3) - P(D_1) - P(D_2) - P(D_3).$$

This causes some overlapping regions to be subtracted more than once (Figure 5.20D). The oval-shaped regions are subtracted twice; the central region is subtracted three times.

Finally, we add all regions contained within exactly two circles back on, so that only the central (orthant) region remains, as Figure 5.20E shows, yielding an equation:

$$\begin{aligned} P(D_1D_2D_3) \;=\; & 1 - P(\bar{D}_1\bar{D}_2\bar{D}_3) \\ & - P(D_1) - P(D_2) - P(D_3) \\ & + P(D_1D_2) + P(D_1D_3) + P(D_2D_3). \end{aligned} \tag{5.97}$$

[1] Note that $\bar{D}_1\bar{D}_2\bar{D}_3$ and $\overline{D_1D_2D_3}$ have distinct meanings. Informally, the former reads, "all the samples are negative," whilst the latter reads, "it is not true that all the samples are positive or zero."

In the case of a stationary zero-mean Gaussian process, the probability of observing three negative samples is equal to that of observing three non-negative samples, so we obtain

$$2P(D_1D_2D_3) = 1 - P(D_1) - P(D_2) - P(D_3) + P(D_1D_2) + P(D_1D_3) + P(D_2D_3). \tag{5.98}$$

Hence, in (5.98), we have successfully reduced the three-dimensional orthant probability to a sum of two- and one-dimensional orthant probabilities.

The General Case

In view of (5.98), one might expect that a d-dimensional orthant probability can be computed recursively. However, as Kedem (1980), referring to David (1953), points out, this is not the case. A general formula, due to Boole[1], exists for relating orthant probabilities of differing dimension:

$$P(\bar{D}_1\bar{D}_2\ldots\bar{D}_d) \equiv 1 - \sum_{0<i\leq d} P(D_i) + \sum_{0<i<j\leq d} P(D_iD_j) - \sum_{0<i<j<k\leq d} P(D_iD_jD_k) + \ldots + (-1)^d P(D_iD_j\ldots D_d). \tag{5.99}$$

The identity (5.99) may be derived in the same way as the three-dimensional case, *viz.*, by conceptually adding and subtracting regions in a Venn diagram. Note that (5.98) is a special case of (5.99) when $d = 3$.

David (1953) dedicated a short paper to a property of (5.99), which is relevant to our present enquiry. Let us grant that $P(D_1\ldots D_N) = P(\bar{D}_1\ldots\bar{D}_d)$. If d is odd, the d-dimensional orthant probability on the right-hand side is preceded by (-1), so that adding $P(D_1\ldots D_d)$ to both sides causes the orthant probability to double up on the left-hand side, as in (5.98). However, for even d, the orthant term is added to both sides and so cancels via subtraction. As a result, David notes, it is not possible to construct a recurrence relation to reduce even-dimensional orthant probabilities into a sum of lower-dimensional cases.

In light of the above, it is instructive to consider the only even-dimensional orthant probability for which we do possess an expression: the bivariate case. Knowledge of $P(D_1)$ and $P(D_2)$, even with the constraint $P(D_1D_2) = P(\bar{D}_1\bar{D}_2)$, does not provide enough information to determine $P(D_1D_2)$. This is the most trivial example in which an even-dimensional orthant probability fails to yield a recurrence relation; and yet a solution is known, and this suggests a route to finding orthant probabilities that is not based on recursion.

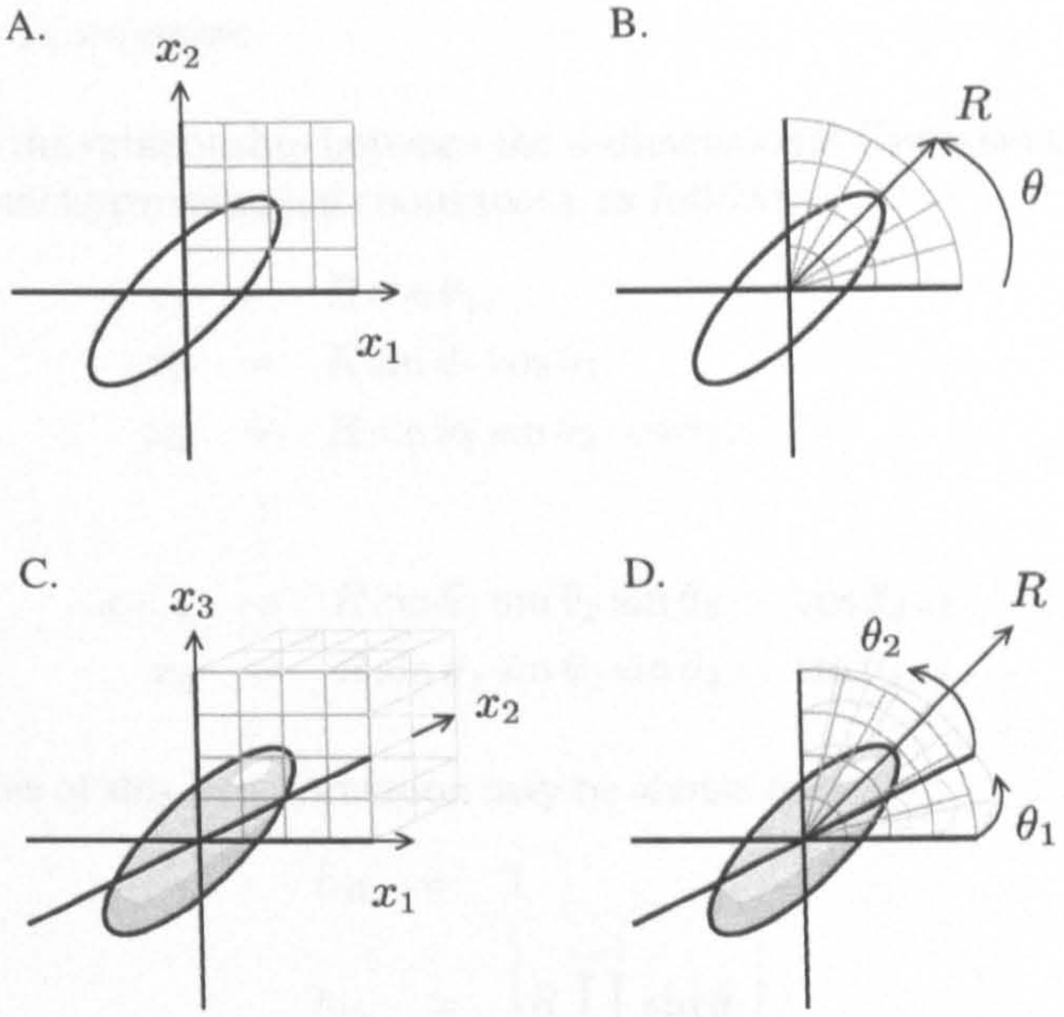


Figure 5.21: A. 2D orthant in Cartesian coordinates. B. 2D orthant in polar coordinates. C. 3D orthant in Cartesian coordinates. D. 3D orthant in spherical coordinates.

5.4.2 Direction Integration

Our next strategy is to directly evaluate (5.96) by changing variables and performing the integration in a hyperspherical coordinate system—a natural extension of polar (2D) and spherical (3D) coordinates into higher dimensions. To illustrate this idea, Figure 5.21A shows a contour of a two-dimensional Gaussian density function and sketches the Cartesian coordinate lines covering the region of integration. Figure 5.21B shows how an equivalent integration can be performed in polar coordinates, where an angle, θ_1, spans one quarter of the plane, and a radius, R, ranges from 0 to $+\infty$. Similarly, the three-dimensional orthant region can be integrated with reference to Cartesian coordinates (Figure 5.21C) or spherical coordinates (Figure 5.21D). To reach every point within a d-dimensional polar coordinate system, a radius and $d-1$ angular coordinates are needed.

[1] A slightly modified version of (5.99) tends to be attributed to Boole (David, 1953). Kedem (1980, p34) refers to the result as "Boole's formula". It can be derived using binomial coefficients (or Pascal's triangle).

Hyperspherical Coordinates

We shall define the relationship between the d-dimensional Cartesian coordinates and the d-dimensional hyperspherical coordinates, as follows:

$$\begin{aligned} x_1 &= R\cos\theta_1, \\ x_2 &= R\sin\theta_1\cos\theta_2, \\ x_3 &= R\sin\theta_1\sin\theta_2\cos\theta_3, \\ &\vdots \\ x_{d-1} &= R\sin\theta_1\sin\theta_2\sin\theta_3\cdots\cos\theta_{d-1}, \\ x_d &= R\sin\theta_1\sin\theta_2\sin\theta_3\cdots\sin\theta_{d-1}. \end{aligned} \tag{5.100}$$

The scale factors of this transformation may be shown to be

$$h_R = 1 \tag{5.101}$$

$$h_{\theta_i} = \left| R\prod_{j=1}^{i-1}\sin\theta_j \right|. \tag{5.102}$$

The Jacobian is then given by the product of scale factors,

$$J = \left| h_R\prod_{i=1}^{d-1} h_{\theta_i} \right| = \left| R^{d-1}\sin^{d-2}\theta_1\sin^{d-3}\theta_2\ldots\sin^2\theta_{d-3}\sin\theta_{d-2} \right|. \tag{5.103}$$

The all-positive d-dimensional orthant probability for a Gaussian distribution with zero mean and covariance matrix Σ is found by integrating over the region of the p.d.f. in which all the variables are positive, i.e.,

$$\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\int_0^\infty\cdots\int_0^\infty \exp\left(\frac{\mathbf{x}^T\Sigma^{-1}\mathbf{x}}{-2}\right) dx_1\cdots dx_d. \tag{5.104}$$

Replacing the variables $x_1, x_2, \ldots, x_d$ with the variables $R, \theta_1, \theta_2, \ldots, \theta_{d-1}$, including the Jacobian J, and changing the bounds, this probability can be expressed

$$\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\int_0^{\pi/2}\cdots\int_0^{\pi/2}\int_0^\infty \left|R^{d-1}\sin^{d-2}\theta_1\ldots\sin^2\theta_{d-3}\sin\theta_{d-2}\right| \times \exp\left(\frac{\Theta^T\Sigma^{-1}\Theta R^2}{-2}\right) dR d\theta_1\cdots d\theta_{d-1}, \tag{5.105}$$

where

$$\Theta = \begin{bmatrix} \cos\theta_1 \\ \sin\theta_1\cos\theta_2 \\ \vdots \\ \sin\theta_1\sin\theta_2\sin\theta_3\cdots\cos\theta_{d-1} \\ \sin\theta_1\sin\theta_2\sin\theta_3\cdots\sin\theta_{d-1} \end{bmatrix}. \tag{5.106}$$

Using the following general solution for odd n, found by repeated integration,

$$\int_0^\infty R^n \exp \frac{aR^2}{-2} dR = \frac{1}{2}\left(\frac{n-1}{2}\right)!\left(\frac{2}{a}\right)^{\frac{n+1}{2}}, \tag{5.107}$$

we can solve the innermost integral in (5.105), for even d, leaving

$$\frac{(d/2-1)!}{2\pi^{d/2}|\Sigma|^{1/2}} \int_0^{\pi/2} \cdots \int_0^{\pi/2} \frac{\sin^{d-2}\theta_1 \ldots \sin^2\theta_{d-3} \sin\theta_{d-2}}{(\Theta^T \Sigma^{-1} \Theta)^{d/2}} d\theta_1 \cdots d\theta_{d-1}. \tag{5.108}$$

The expression for the orthant probability given in (5.108) is a ratio of polynomials in $\exp i\theta_1, \exp i\theta_1, \ldots, \exp i\theta_{d-1}$. A general analytical solution for this kind of integral does not appear to be reported in the literature, and (5.108) seems as insoluble as the original multiple integral given in (5.96). In closing, however, we may note two apparent advantages of the hyperspherical approach: i) the order of the multiple integral has been reduced from d to $d-1$, and ii) the bounds on every integral are finite rather than infinite, which may facilitate a numerical integration approach.

5.4.3 An Exact Solution from Geometry: 2D and 3D

The direct integration approach attempted above faces a difficult integrand (namely, the multivariate Gaussian p.d.f.), but a simple region of integration (i.e., the positive orthant). We shall now tackle the problem from a geometric perspective, and in so doing, simplify the integrand at the expense of complicating the region of integration. This is analogous to solving $\int_0^4 (x+1)dx$ by finding the area enclosed by the lines $x=0$, $x=4$, $y=0$ and $y-x-1=0$, as opposed to using antiderivatives.

The integration we are looking to perform is

$$P(x_1 \geq 0, \ldots, x_d \geq 0) = \frac{1}{(2\pi)^{d/2}|\Sigma_{\mathcal{X}}|^{1/2}} \int_0^\infty \cdots \int_0^\infty \exp\left(\frac{\mathbf{x}^T \Sigma_{\mathcal{X}}^{-1} \mathbf{x}}{-2}\right) dx_1 \cdots dx_d. \tag{5.109}$$

The kind of procedure we are about to employ should by now be familiar to the reader from earlier sections. We first replace the variables $\mathbf{x}^T \triangleq (x_1, \ldots, x_d)^T$ with another set of uncorrelated and normalised variables $\mathbf{y}^T \triangleq (y_1, \ldots, y_d)^T$ using

$$\mathbf{y} = T\mathbf{x}, \tag{5.110}$$

where $T = U\sqrt{D^{-1}}$, and U and D are the matrix of eigenvectors (in columns) and eigenvalues (along the diagonal) of $\Sigma_{\mathcal{X}}^{-1}$, respectively. $\Sigma_{\mathcal{Y}}^{-1}$ is the $d \times d$ identity matrix. Performing the change of variables in (5.109) gives

$$P(x_1 \geq 0, \ldots, x_d \geq 0) = \frac{|T|}{(2\pi)^{d/2}|\Sigma_{\mathcal{X}}|^{1/2}} \int_{\mathcal{R}'_+} \exp\left(\frac{\mathbf{y}^T \mathbf{y}}{-2}\right) dy_1 \cdots dy_d, \tag{5.111}$$

where new region of integration is denoted $\mathcal{R}'_+$.

Three important results can be combined to our advantage. First, the new integrand is unaffected by rotations about the origin. To show this briefly, if we assume that a vector $\mathbf{y}_2$ has been obtained by rotating an arbitrary vector, $\mathbf{y}_1$, around the origin using the rotation matrix T_R, i.e.,

$$\mathbf{y}_2 = T_R \mathbf{y}_1,$$

then, because a rotation matrix is orthonormal,

$$\mathbf{y}_2^T \mathbf{y}_2 = (T_R \mathbf{y}_1)^T (T_R \mathbf{y}_1) = \mathbf{y}_1^T (T_R^T T_R) \mathbf{y}_1 = \mathbf{y}_1^T \mathbf{y}_1. \tag{5.112}$$

Second, the region of integration is a cone formed by d half-lines, which converge at the origin and extend outwards infinitely. The unit vectors that point along the edges of this cone are the column vectors of T^{-1} rescaled to unit magnitude. (The region is said to be *subtended* by the vectors.)

Third, the directions in which the unit eigenvectors point do not affect the angles between the column vectors of T^{-1}, which is essential, because the behaviour of eigenvalue solvers such as `eig` in MATLAB do not specify the direction of the output eigenvectors.

Two-dimensional Example

Suppose we wish to compute the probability that two consecutive samples, x_1 and x_2, of a zero mean wide-sense stationary Gaussian process are both positive. If it is known (e.g.) that $\rho[1]$ is 0.9, then the correlation matrix governing the pair of samples is

$$\Sigma_X = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}.$$

When the samples are correlated, a contour of the probability density function appears oval-shaped, as in Figure 5.22A. The region of integration is subtended by the column vectors of the 2×2 identity matrix. By appropriately reshaping the integrand, we can effectively transfer the difficulty of the problem into the region of integration instead.

The linear transform which correlates the samples is associated to the matrix

$$T^{-1} = \begin{pmatrix} -0.5130 & -0.5130 \\ -2.2361 & 2.2361 \end{pmatrix}.$$

Multiplying the identity matrix (i.e., our spanning vectors) by T^{-1} simply leaves T^{-1}. Figure 5.22B shows how, although the integrand is now radially uniform, its contours having been stretched from ovals into circles. The column vectors of T^{-1} now enclose the transformed orthant region.

We normalise the columns of T^{-1} to unit length and place them into a new matrix,

$$V = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{pmatrix} = \begin{pmatrix} -0.2236 & -0.2236 \\ -0.9747 & 0.9747 \end{pmatrix}.$$

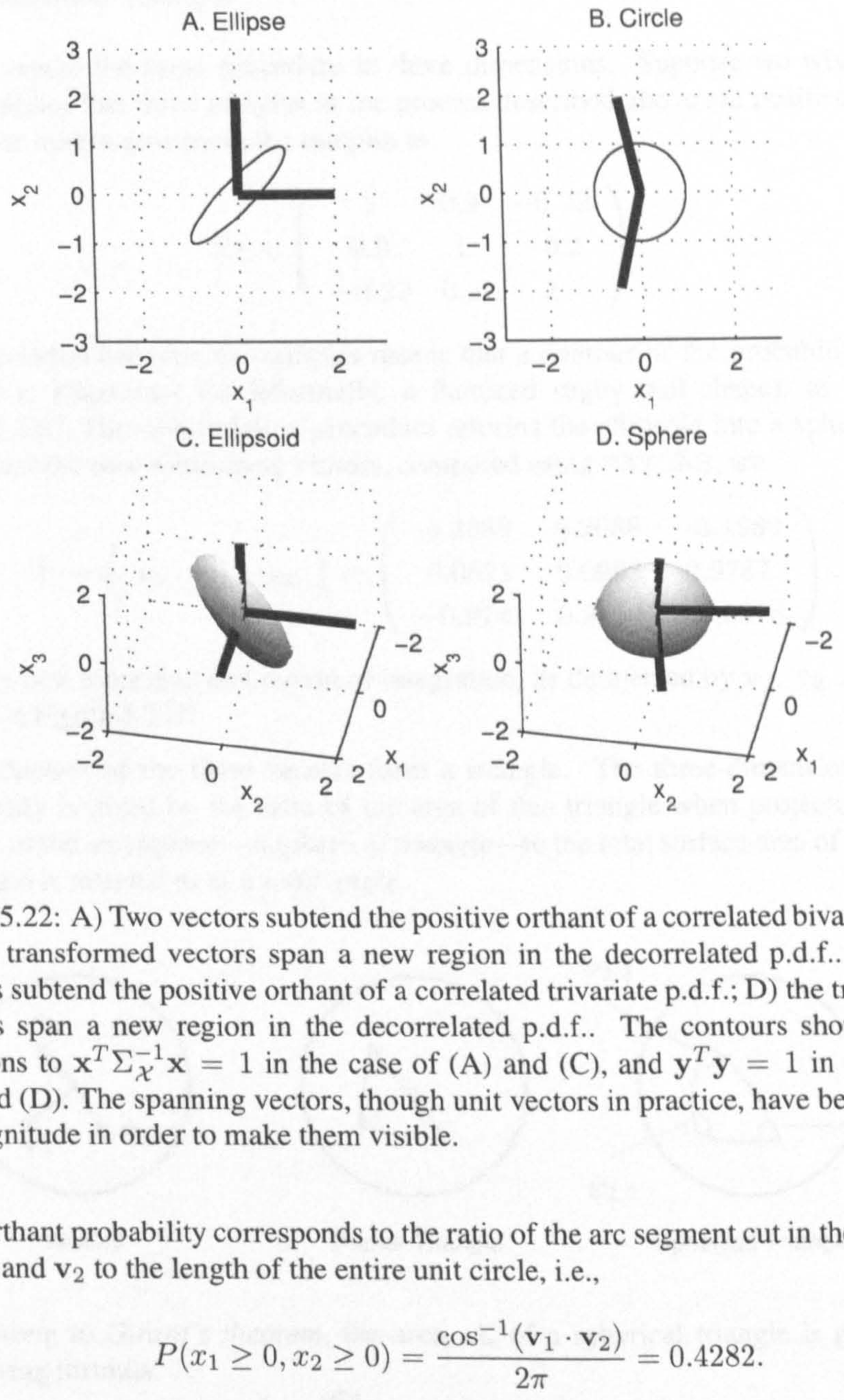


Figure 5.22: A) Two vectors subtend the positive orthant of a correlated bivariate p.d.f.; B) the transformed vectors span a new region in the decorrelated p.d.f.. C) Three vectors subtend the positive orthant of a correlated trivariate p.d.f.; D) the transformed vectors span a new region in the decorrelated p.d.f.. The contours shown are the solutions to $\mathbf{x}^T \Sigma_{\mathcal{X}}^{-1} \mathbf{x} = 1$ in the case of (A) and (C), and $\mathbf{y}^T \mathbf{y} = 1$ in the case of (B) and (D). The spanning vectors, though unit vectors in practice, have been doubled in magnitude in order to make them visible.

The orthant probability corresponds to the ratio of the arc segment cut in the unit circle by $\mathbf{v}_1$ and $\mathbf{v}_2$ to the length of the entire unit circle, i.e.,

$$P(x_1 \geq 0, x_2 \geq 0) = \frac{\cos^{-1}(\mathbf{v}_1 \cdot \mathbf{v}_2)}{2\pi} = 0.4282.$$

The probability of a zero crossing is defined as

$$\begin{aligned} 2P(x_1 \geq 0, x_2 < 0) &= 2\left[P(x_1 \geq 0) - P(x_1 \geq 0, x_2 \geq 0)\right] \\ &= 0.1436. \end{aligned}$$

This is identical to the result found using (1.12).

Three-dimensional Example

We now repeat the same procedure in three dimensions. Suppose we wish to find the probability that three samples in the process described above are positive, and the correlation matrix governing the samples is

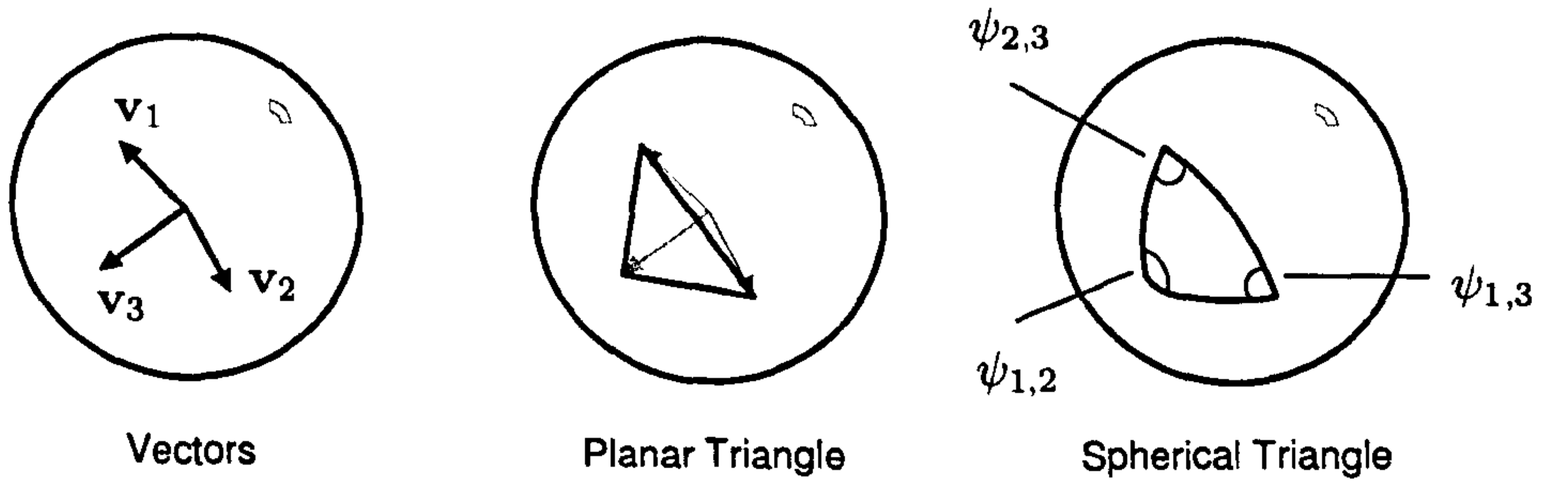
$$\Sigma_3 = \begin{pmatrix} 1 & 0.9 & -0.22 \\ 0.9 & 1 & -0.2 \\ -0.22 & 0.2 & 1 \end{pmatrix}.$$

The correlation between the samples means that a contour of the probability density function is *ellipsoidal* (or informally, a flattened rugby ball shape), as shown in Figure 5.22C. The decorrelation procedure reforms the ellipsoid into a sphere of unit radius; and the new subtending vectors, computed using MATLAB, are

$$V = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{pmatrix} = \begin{pmatrix} 0.2089 & 0.2088 & -0.1989 \\ 0.0871 & 0.0983 & 0.9787 \\ -0.9741 & 0.9730 & -0.0513 \end{pmatrix}.$$

Both the new integrand and region of integration, as delineated by $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$, are plotted in Figure 5.22D.

The endpoints of the three vectors form a triangle. The three-dimensional orthant probability is given by the ratio of the area of this triangle when projected onto the surface of the unit sphere—a *spherical triangle*—to the total surface area of the sphere. This ratio is referred to as a *solid angle*.



According to *Girard's theorem*, the area, A, of a spherical triangle is given by the following formula:

$$A = R^2(\psi_{1,2} + \psi_{2,3} + \psi_{1,3} - \pi), \tag{5.113}$$

where $\psi_{1,2}$, $\psi_{2,3}$ and $\psi_{1,3}$ are the interior angles of the triangle, in radians, as it appears inscribed on the sphere's surface, and R is the radius of the sphere (i.e., $R = 1$). These angles are found to be

$$\begin{aligned} \psi_{1,2} &= 2.6906 \\ \psi_{2,3} &= 1.3694 \\ \psi_{1,3} &= 1.3490. \end{aligned}$$

The orthant probability is then determined:

$$P(x_1 \geq 0, x_2 \geq 0, x_3 \geq 0) = \frac{2.6906 + 1.3694 + 1.3490 - \pi}{4\pi} = 0.1804.$$

This result is identical to that given by the recurrence relation in (5.98) when applied to two-dimensional orthant probabilities.

5.4.4 An Exact Solution from Geometry: The General Case

The geometric approach taken in the two- and three-dimensional cases have proven to be successful, and it is natural to pursue next the question of whether the procedure extends to higher dimensions. It is sufficient for our purposes to develop, or adapt from the geometry literature, a subset of simple geometric principles that allow us to *intuitively* define and manipulate the unseen, high-dimensional analogues of spheres and triangles in a manner that augments the lower-dimensional examples above.

There are two separate considerations:

1. Does the notion of a line projected onto the perimeter of a circle, or a triangle projected onto the surface of a sphere continue into higher dimensions? Is this the shape that results when the decorrelating transformation (T^{-1}) is applied to the all-positive orthant?

2. If such a shape can be identified, does the ratio of its hyper-area to the total hyperspherical surface express to the orthant probability? If so, can both areas be evaluated in order to find the orthant probability?

To address these questions, we shall start with a series of definitions.

Definitions

Definition 1. *The* d-dimensional Euclidean space, $\mathbb{E}^d$, *is defined as the set of all ordered* d*-tuples whose components are real numbers, i.e.,*

$$\mathbb{E}^d = \{(p_1, p_2, \ldots p_d) : p_1, p_2, \ldots, p_d \in \mathbb{R}\}.$$

An element of $\mathbb{E}^d$ *is called a* point. *We shall adopt the convention of identifying a point* P *with the column vector* $\overrightarrow{OP} = \mathbf{p} = (p_1, p_2, \ldots, p_d)^T$. *The* Euclidean distance *between two points,* $\mathbf{p}_1$ *and* $\mathbf{p}_2$*, is defined as*

$$\text{dist}(\mathbf{p}_1, \mathbf{p}_2) = \sqrt{(\mathbf{p}_1 - \mathbf{p}_2)^T(\mathbf{p}_1 - \mathbf{p}_2)}.$$

Definition 2. *The* d-dimensional (Euclidean) positive orthant, $\mathbb{O}^d \subset \mathbb{E}^d$, *is defined as the set of all ordered* d*-tuples whose components are all non-negative real numbers.*

Definition 3. *We define the* unit n-sphere *as*

$$\mathcal{S}^n = \{\mathbf{p} \in \mathbb{E}^{n+1} : |\mathbf{p}| = 1\}.$$

Note that n refers to the dimensionality of the *boundary* of the shape. For example, a circle is a 1-sphere, and the usual sphere in three dimensions is a 2-sphere. The points on the interior of the n-sphere are not members.

Many shapes can be constructed by specifying the parameters of a set comprehension. The definitions often include a predicate to check whether a certain parameterisation results in a valid construction.

Definition 4. *Let* $[\mathbf{p}_1, \mathbf{p}_2]$ *denote the closed line segment between* $\mathbf{p}_1, \mathbf{p}_2 \in \mathbb{E}^d$.

Definition 5. *Let* $\mathbf{p} \in \mathbb{E}^n$, $\mathbf{p} \neq 0$. *We define a* half-line *as*

$$\text{halfline}(\mathbf{p}) = \{\alpha\mathbf{p} : \alpha \in \mathbb{R}, \alpha > 0\}.$$

Line segments and half-lines are one-dimensional, regardless of the dimensionality of the space in which they are embedded.

Definition 6. *Let* $\mathcal{P}$ *denote a set of* $n{+}1$ *points,* $\mathbf{p}_1, \ldots, \mathbf{p}_{n+1} \in \mathbb{E}^d, d \geq n$, *such that* $\mathcal{P}$ *is not a subset of a* $(n{-}1)$*-hyperplane. We define the* n-simplex *for points (*vertices*)* $\mathcal{P}$ *as follows.*

$$\text{simplex}(\mathcal{P}) = \begin{cases} \mathbf{p}_1 & n = 0 \\ [\mathbf{p}_1, \mathbf{p}_2] & n = 1 \\ \{\mathbf{q} \in [\mathbf{l}_1, \mathbf{l}_2] : \mathbf{l}_1, \mathbf{l}_2 \in \text{bd}(\mathcal{T})\} & n > 1. \end{cases}$$

where we define the boundary, $\text{bd}(\mathcal{T})$, *of a simplex* $\mathcal{T}$ *with vertices* $\mathcal{P}$ *as follows.*

$$\text{bd}(\mathcal{T}) = \{\mathbf{q} \in \text{simplex}(\mathcal{Q}) : \mathcal{Q} \subset \mathcal{P}, |\mathcal{Q}| = |\mathcal{P}| - 1\}.$$

Definition 7. *Let* $\mathcal{P}$ *denote a set of* $n{+}1$ *points,* $\mathbf{p}_1, \ldots, \mathbf{p}_{n+1} \in \mathbb{E}^d$, $d \geq n$, *such that* $\mathcal{P}$ *is not a subset of a* $(n{-}1)$*-hyperplane, and* $0 \notin \text{simplex}(\mathcal{P})$, *and no two vertices* $\mathbf{p}_1, \mathbf{p}_2 \in \mathcal{P}$ *are members of the same halfline. Let* $\mathcal{T} = \text{simplex}(\mathcal{P})$. *We define the* n-cone *specified by* $\mathcal{T}$ *as*

$$\text{cone}(\mathcal{T}) = \{h \in \text{halfline}(\mathbf{q}) : \mathbf{q} \in \mathcal{T}\}.$$

Furthermore, we define the boundary of the n*-cone specified by* $\mathcal{T}$ *as*

$$\text{bd}(\text{cone}(\mathcal{T})) = \{h \in \text{halfline}(\mathbf{q}) : \mathbf{q} \in \text{bd}(\mathcal{T})\}.$$

The analogue of spherical triangle is now described simply by a set intersection.

Definition 8. *Let* $\mathcal{P}$ *denote a set of* $n{+}1$ *points,* $\mathbf{p}_1, \ldots, \mathbf{p}_{n+1} \in \mathbb{E}^d$, $d \geq n$, *such that* $\mathcal{P}$ *is not a subset of a* $(n{-}1)$*-hyperplane, and* $0 \notin \text{simplex}(\mathcal{P})$, *and no two vertices* $\mathbf{p}_1, \mathbf{p}_2 \in \mathcal{P}$ *are members of the same halfline. Let* $\mathcal{T} = \text{simplex}(\mathcal{P})$. *We define the* unit n-spherical simplex *with vertices* $\mathcal{P}$ *as*

$$\text{sphsimplex}(\mathcal{T}) = \text{cone}(\mathcal{T}) \cap \mathcal{S}^n,$$

and its boundary as

$$\text{bd}(\text{sphsimplex}(\mathcal{T})) = \text{bd}(\text{cone}(\mathcal{T})) \cap \mathcal{S}^n.$$

Definition 9. *Let Q denote a linear transformation associated to matrix $\mathbf{Q}$. Let $\mathcal{S} \subset \mathbb{E}^n$ denote a space (e.g., a simplex or half-line). Let $\mathbf{p} \in \mathbb{E}^d$. The application of Q to $\mathbf{p}$ has the following definition.*

$$Q(\mathbf{p}) = \mathbf{Q}\,\mathbf{p}.$$

The application of Q to $\mathcal{S}$ is similarly defined.

$$Q(\mathcal{S}) = \{Q(\mathbf{q}) : \mathbf{q} \in \mathcal{S}\}.$$

Propositions

Proposition 1. *Let Q denote an invertible linear transformation. If $\mathbf{p} \in \mathbb{E}^d$ and $\mathbf{p} \neq 0$ then $Q(\mathbf{p}) \neq 0$.*

Proposition 2. *Let $\mathcal{L} = [\mathbf{l}_1, \mathbf{l}_2]$, $\mathbf{l}_1 \neq \mathbf{l}_2$, and let Q denote an invertible linear transformation. It follows that*

$$\mathbf{p} \in \mathcal{L} \leftrightarrow Q(\mathbf{p}) \in Q(\mathcal{L}).$$

Proposition 3. *Let Q denote an invertible linear transformation and $\mathcal{P}$ denote a set of points $\{\mathbf{p}_1, \ldots, \mathbf{p}_{d+1}\}$. That no two members of $\mathcal{P}$ are members of the same halfline implies that no two members of $Q(\mathcal{P})$ are members of the same halfline.*

Proposition 4. *Let $\mathcal{C}$ denote a cone. For any constant, $\alpha > 0$,*

$$\mathbf{p} \in \mathcal{C} \leftrightarrow \alpha\mathbf{p} \in \mathcal{C}.$$

Proposition 5. *Let $\mathcal{P}$ denote a set of points in $\mathbb{E}^d$ for which exactly one coordinate is one and the rest are zero. It follows that*

$$\text{cone}(\text{simplex}(\mathcal{P})) = \mathbb{O}^d \setminus \{0\}.$$

Theorems

Lemma 1. *Let $\mathcal{T}$ denote an n-simplex, such that $\mathcal{T} \subset \mathbb{E}^d$, $d \geq n$. Choose a point $\mathbf{p} \in \mathbb{E}^d$. Let Q denote an invertible linear transformation. The following must hold.*

$$\mathbf{p} \in \mathcal{T} \rightarrow Q(\mathbf{p}) \in Q(\mathcal{T}).$$

Proof. For $n = 1$, this theorem is proven by Proposition 2.

For $n > 1$, assume the theorem holds for $n - 1$.

By Definition 6, $\mathbf{p} \in \mathcal{T}$ implies the existence of a line segment $\mathcal{L} = [\mathbf{l}_1, \mathbf{l}_2]$, such that $\mathbf{l}_1, \mathbf{l}_2 \in \text{bd}(\mathcal{T})$. Let $\mathbf{l}'_1 = Q(\mathbf{l}_1)$ and $\mathbf{l}'_2 = Q(\mathbf{l}_2)$ and $\mathcal{T}' = Q(\mathcal{T})$.

From Definition 6, $\text{bd}(\mathcal{T})$ is a set of $(n-1)$-simplices. Using the inductive hypothesis and Definition 9, $\mathbf{l}_1, \mathbf{l}_2 \in \text{bd}(\mathcal{T})$ implies that $\mathbf{l}'_1, \mathbf{l}'_2 \in \text{bd}(\mathcal{T}')$.

Let $\mathbf{p}' = Q(\mathbf{p})$. By Proposition 2, $\mathbf{p} \in \mathcal{L}$ implies that $\mathbf{p}' \in [\mathbf{l}'_1, \mathbf{l}'_2]$. By Definition 6, the existence of such a line segment with end points in $\text{bd}(\mathcal{T}')$ implies that $\mathbf{p}' \in \mathcal{T}'$. □

Lemma 2. *Let $\mathcal{T}$ denote an n-simplex, such that $\mathcal{T} \subset \mathbb{E}^d$, $d \geq n$. Choose a point $\mathbf{p} \in \mathbb{E}^d$. Let Q denote an invertible linear transformation. The following must hold.*

$$\mathbf{p} \notin \mathcal{T} \rightarrow Q(\mathbf{p}) \notin Q(\mathcal{T}).$$

Proof. Let Q^{-1} refer to the inverse of Q. From Lemma 1,

$$Q(\mathbf{p}) \in Q(\mathcal{T}) \rightarrow Q^{-1}(Q(\mathbf{p})) \in Q^{-1}(Q(\mathcal{T}));$$

thus, $Q(\mathbf{p}) \in Q(\mathcal{T}) \rightarrow \mathbf{p} \in \mathcal{T}$. □

Lemma 3. *Let $\mathcal{P}$ denote a set of $n+1$ points, $\mathbf{p}_1, \ldots, \mathbf{p}_{n+1} \in \mathbb{E}^d$, $d \geq n$, such that $\mathcal{P}$ is not a subset of a $(n-1)$-hyperplane, and $0 \notin \text{simplex}(\mathcal{P})$, and no two vertices $\mathbf{p}_1, \mathbf{p}_2 \in \mathcal{P}$ are members of the same halfline. Let $\mathcal{T} = \text{simplex}(\mathcal{P})$. Let Q denote an invertible linear transformation. The following must hold.*

$$Q(\text{cone}(\mathcal{T})) = \text{cone}(Q(\mathcal{T})).$$

Proof. The following is a direct proof.

From Definitions 7 and 9,

$$\begin{aligned} Q(\mathcal{C}) &= Q(\{h \in \text{halfline}(\mathbf{q}) : \mathbf{q} \in \mathcal{T}\}) \\ &= \{h \in Q(\text{halfline}(\mathbf{q})) : \mathbf{q} \in \mathcal{T}\}. \end{aligned}$$

From Definition 5, Proposition 1, and Lemmas 1 and 2, along with the fact that Q is linear,

$$\begin{aligned} Q(\mathcal{C}) &= \{h \in Q(\{\alpha\mathbf{q} : \alpha \in \mathbb{R}, \alpha > 0\}) : \mathbf{q} \in \mathcal{T}\} \\ &= \{h \in (\{\alpha Q(\mathbf{q}) : \alpha \in \mathbb{R}, \alpha > 0\}) : \mathbf{q} \in \mathcal{T}\} \\ &= \{h \in (\{\alpha\mathbf{q} : \alpha \in \mathbb{R}, \alpha > 0\}) : Q(\mathbf{q}) \in \mathcal{T}\} \\ &= \{h \in (\{\alpha\mathbf{q} : \alpha \in \mathbb{R}, \alpha > 0\}) : \mathbf{q} \in Q(\mathcal{T})\} \\ &= \text{cone}(Q(\mathcal{T})). \end{aligned}$$

□

Theorem 1. *Let Q denote an invertible linear transformation. There always exists a simplex $\mathcal{T}$, such that*

$$\{\mathbf{q}/|\mathbf{q}| : \mathbf{q} \in Q(\mathbb{O}^d), \mathbf{q} \neq 0\} = \text{sphsimplex}(\mathcal{T}).$$

Proof. From Proposition 5 and Lemma 3,

$$\begin{aligned} &\{\mathbf{q}/|\mathbf{q}| : \mathbf{q} \in Q(\text{cone}(\text{simplex}(\mathcal{P})))\} \\ &= \{\mathbf{q}/|\mathbf{q}| : \mathbf{q} \in \text{cone}(Q(\text{simplex}(\mathcal{P})))\} \\ &= \{\mathbf{q}/|\mathbf{q}| : \mathbf{q} \in \text{cone}(\mathcal{T}))\} \end{aligned}$$

where $\mathcal{T} = Q(\text{simplex}(\mathcal{P}))$ and $\mathcal{T}$ is a $(d-1)$-simplex (by Lemmas 1 and 2). By Proposition 5, $\mathcal{P}$ is a set of points in $\mathbb{E}^d$, for which exactly one coordinate is one and the rest are zero.

From Definition 8, $\text{sphsimplex}(\mathcal{T}) = \text{cone}(\mathcal{T}) \cap \mathcal{S}^{d-1}$.

From Definition 3, $\mathbf{q}/|\mathbf{q}| \in \mathcal{S}^{d-1}$ because $|\mathbf{q}/|\mathbf{q}|| = 1$ and $\mathbf{q}/|\mathbf{q}| \in \mathbb{E}^d$.

From Proposition 4, noting that $|\mathbf{q}|$ is a constant satisifying $|\mathbf{q}| > 0$,

$$\mathbf{p} \in \{\mathbf{q}/|\mathbf{q}| : \mathbf{q} \in \text{cone}(\mathcal{T})\} \leftrightarrow \mathbf{p} \in \text{cone}(\mathcal{T}).$$

In consequence, $\{\mathbf{q}/|\mathbf{q}| : \mathbf{q} \in Q(\mathbb{O}^d), \mathbf{q} \neq 0\} = \text{sphsimplex}(\mathcal{T})$. □

Theorem 2. *Adapted from Muller (1959); see also Marsaglia (1972). If $Y_1, \ldots, Y_d$ are standard normal variates, put $S = Y_1^2 + \ldots + Y_d^2$ and form*

$$\mathbf{p} = \left(\frac{Y_1}{\sqrt{S}}, \ldots, \frac{Y_d}{\sqrt{S}} \right)^T.$$

The random vector $\mathbf{p}$ is uniformly distributed over the surface of the $(d-1)$-sphere.

Closing Remarks

From Theorem 1, the probability that the non-zero random vector $\mathbf{x}$ in $\mathbb{E}^d$ is a member of the positive orthant is equal to the probability that the random vector $\mathbf{y}/|\mathbf{y}|$ is a member of the spherical-simplicial region on the $(d-1)$-sphere, because $\mathbf{y}$ is obtained from $\mathbf{x}$ via an invertible linear transformation, T.

The first of the two questions set out at the start of this section has been answered conclusively. An analogous relationship does exist between the triangle and sphere in three dimensions, and the $(d-1)$-simplex and $(d-1)$-sphere in d dimensions. It remains to address the second question.

From Theorem 2, we see that $\mathbf{y}/|\mathbf{y}|$ is distributed uniformly over the surface of the $(d-1)$-sphere, because the elements of $\mathbf{y}$ are Gaussian with zero mean, zero covariance and unit variance. Consequently, the ratio of the *content* (i.e., hypervolume) of the spherical-simplex to that of the entire hyperspherical surface is equal to the orthant probability. This ratio is the high-dimensional analogue of a solid angle.

At the present time, the solution for the general solid angle in $\mathbb{E}^d$ is elusive for $d > 3$. Ribando (2006) comments, "there appears to be no closed form expression for the measure of an n-dimensional solid angle for $n > 3$". This agrees with the earlier assessment of David (1953) that "an exact expression for this solid angle appears to be known for three dimensions only". Concerning $d = 4$, Abrahamson (1964) asserts that "the orthant probability, i.e., the probability that all the [variables] will be simultaneously positive, is not, in general, given by a closed expression."

Although an exact analytical solution is unavailable, several workers have suggested approximations, including series expansions (Ribando, 2006) and decomposition into *orthoschemes* (Abrahamson, 1964). Interestingly, Hajja and Walker (2002) tackle the

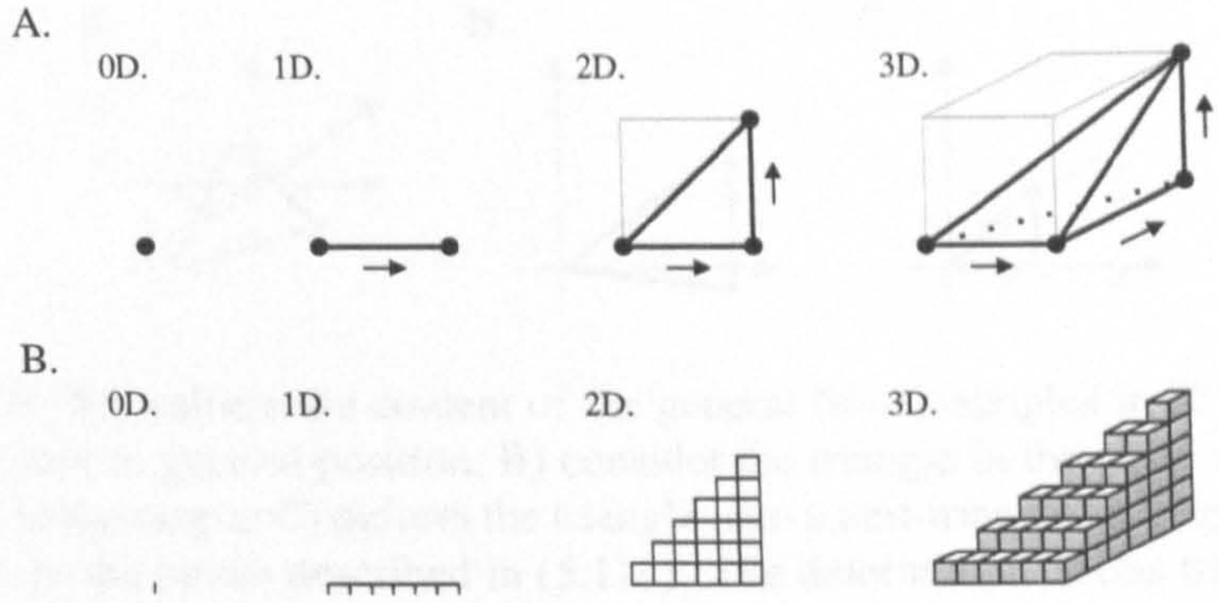
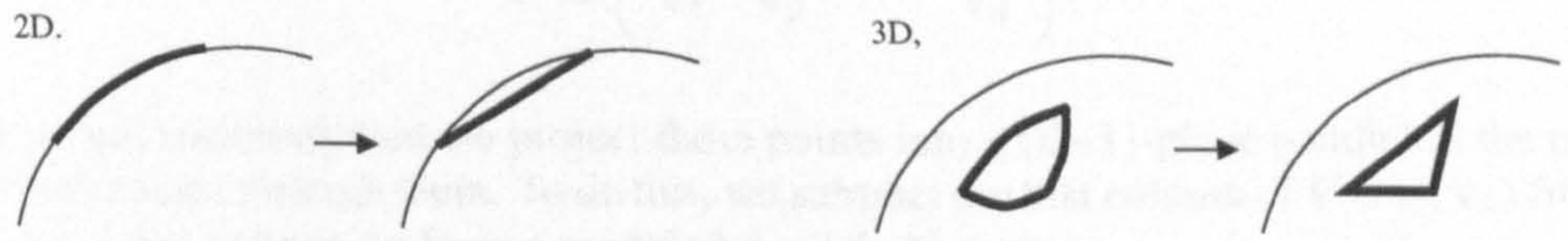


Figure 5.23: Conceptual illustration of how A) a unit d-simplex is constructed and B) its content is evaluated, in 0, 1, 2 and 3 dimensions.

four-dimensional solid angle problem by re-expressing it as a multiple integral similar to the one which appears in (5.108) and solving it numerically. In light of the lack of a generally-approved approach to finding the solid angles in high dimensions, we shall examine the possibility of approximating the content of the $(d-1)$-spherical simplex in $\mathbb{E}^d$ by decomposing it into simpler regions that can be readily integrated.

5.4.5 An Approximate Solution from Geometry

If the unit d-spherical simplex is small, in the sense that its vertices are relatively closely-spaced on the d-sphere, then it may be acceptable to approximate its content by measuring the content of the (hyper-)planar simplex from which it is projected.



Content of the d-Simplex

We shall first consider the content of a simplex formed by joining the vertices of a unit hypercube. The vertex set of such a $(d-1)$-simplex can be generated by starting at the origin and moving a unit distance into each orthogonal dimension. The vertices then correspond to the column vectors of a $(d-1) \times (d-1)$ upper-triangular matrix of ones, i.e.,

$$U = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}. \tag{5.114}$$

A graphical depiction of this process is provided in Figure 5.23A.

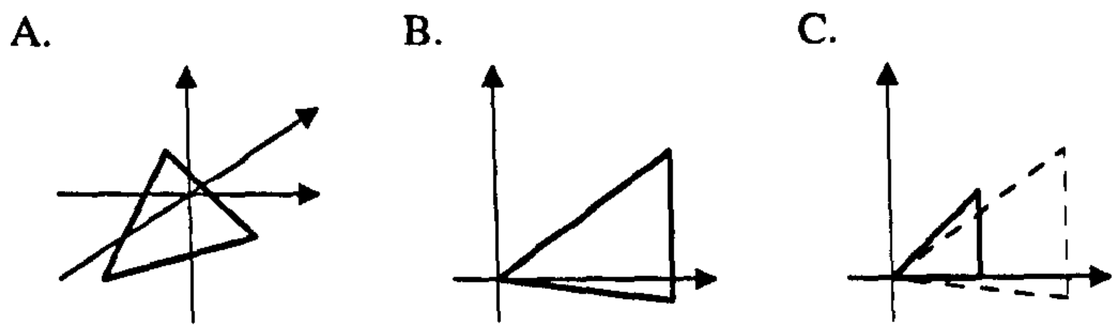


Figure 5.24: To evaluate the content of the general $(d-1)$-simplex in $\mathbb{E}^d$, for $d = 3$: A) the simplex in general position, B) consider the triangle in the plane and translate one vertex to the origin; C) deform the triangle onto a unit-triangle by mapping each of its vertices to the points described in (5.114). The determinant of this transformation relates the change in area.

The content of the (closure of the) $(d-1)$-simplex is found by integration.

$$\int_0^1 \int_0^{x_1} \cdots \int_0^{x_{d-1}} dx_d \, dx_{d-1} \ldots dx_1 = \frac{1}{d!}. \tag{5.115}$$

The rationale behind this integral is conveyed in Figure 5.23B, which illustrates how the closure of a simplex is divided into small hypercubes.

Having determined the content of a unit $(d-1)$-simplex, it remains to discover the content of an arbitrary $(d-1)$-simplex, embedded in a d-dimensional space. To achieve this, we will *deform* the arbitrary simplex onto a unit simplex and allow the determinant of the transformation to inform us of the change in content. Let the vectors corresponding to the vertices of the $(d-1)$-simplex in $\mathbb{E}^d$ be denoted $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_d$, and form the column matrix V from these vectors, i.e.,

$$V = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_d \end{pmatrix}.$$

It is next necessary that we project these points into a $(d-1)$-plane parallel to the one which passes through them. To do this, we subtract the first column of V (i.e., $\mathbf{v}_1$) from every other column, to form a modified $d \times (d-1)$ matrix,

$$V' = \begin{pmatrix} \mathbf{v}_2 - \mathbf{v}_1 & \mathbf{v}_3 - \mathbf{v}_1 & \cdots & \mathbf{v}_d - \mathbf{v}_1 \end{pmatrix}.$$

Let W denote a $d \times (d-1)$ matrix whose columns are the set of orthonormal vectors that span the same space as the column vectors of V'. The columns of the matrix $W^T V'$ describe the vertices of the figure as it appears in the $(d-1)$-dimensional hyperplane that passes through it. (See Figure 5.24B.)

Let U denote the $(d-1) \times (d-1)$ upper-triangular matrix of ones from (5.114). The linear transform which maps the arbitrary simplex onto a unit simplex is associated to the matrix $W^T V' U^{-1}$. (See Figure 5.24.) Consequently, the content, C, of the simplex is given by

$$C = \frac{1}{d!} \left| \left| W^T V' U^{-1} \right| \right|, \tag{5.116}$$

where $|| \cdot ||$ denotes the magnitude of the determinant.

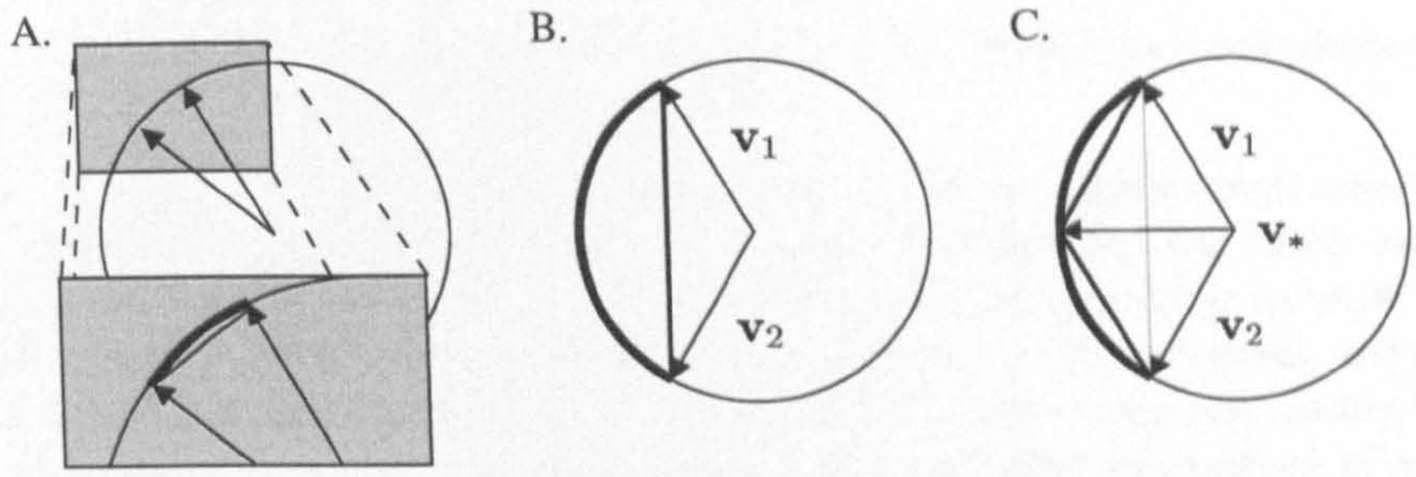


Figure 5.25: A) the length of a line segment constructed under a circular arc is a good approximation to the arc-length, if the angle is small; B) for larger angles, this approximation is very poor; C) the solution is to split the line into two, shorter parts at point $\mathbf{v}_*$, extend these line segments out to the surface, and sum their length. One obtains ever-finer approximations by repeating (C) on the shorter segments; eventually (it is conjectured) this approximation will converge onto the exact solution.

The Cayley-Menger Determinant

The standard approach to evaluating the content of a simplex is the *Cayley-Menger determinant*. First, we define a matrix of squared Euclidean distances,

$$A = \begin{pmatrix} \mathrm{dist}(\mathbf{v}_1, \mathbf{v}_1)^2 & \mathrm{dist}(\mathbf{v}_1, \mathbf{v}_2)^2 & \cdots & \mathrm{dist}(\mathbf{v}_1, \mathbf{v}_d)^2 \\ \mathrm{dist}(\mathbf{v}_2, \mathbf{v}_1)^2 & \mathrm{dist}(\mathbf{v}_2, \mathbf{v}_2)^2 & \cdots & \mathrm{dist}(\mathbf{v}_2, \mathbf{v}_d)^2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{dist}(\mathbf{v}_d, \mathbf{v}_1)^2 & \mathrm{dist}(\mathbf{v}_d, \mathbf{v}_2)^2 & \cdots & \mathrm{dist}(\mathbf{v}_d, \mathbf{v}_d)^2 \end{pmatrix}. \tag{5.117}$$

It is clear that A is symmetric and has zeros along its main diagonal and positive elements everywhere else. Next, B is the $(d{+}1) \times (d{+}1)$ matrix formed by bordering A with ones, such that

$$B = \begin{pmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & A \end{pmatrix}, \tag{5.118}$$

where $\mathbf{1}$ denotes a $d \times 1$ column vector of ones. The content of the $(d{-}1)$-simplex is given by

$$C = \sqrt{\frac{(-1)^d |B|}{2^{d-1}(d-1)!^2}}. \tag{5.119}$$

Approximation with Subdivision

The method presented in the section above approximates, in three dimensions, the surface area of a triangle projected onto the surface of a sphere—that is, a spherical triangle—as the area of the original planar triangle. The method also extends naturally to d dimensions; so, by analogy: it approximates the content of a $(d{-}1)$-simplex projected onto the surface of a $(d{-}1)$-sphere, as the content of the original hyperplanar

simplex. Evidently, this approximation is poor if the vertices are not clustered in a small region of the hypersphere to begin with.

Figures 5.25A and 5.25B respectively show that, in the two dimensional case, a linear approximation is adequate if the vectors are closely-spaced (A), but much too low if they are widely-spaced (B). An obvious solution, stated informally, is to break the line into two shorter, connected lines, by adding a vertex, $\mathbf{v}_*$, to its mid-point, and pushing $\mathbf{v}_*$ out onto the surface of the circle, as Figure 5.25C shows. One can readily imagine that the same idea applies to a planar triangle situated under the surface of a sphere. If the triangle is split into two smaller triangles by introducing a new vertex along one edge, and this vertex is projected onto the spherical surface, then the summed area of the two triangles provides a better approximation to the spherical triangle than the original, planar triangle.

We can extend this principle to any number of dimensions with relative ease. Let $\mathcal{V}$ denote the set of vertices $\mathbf{v}_1, \ldots, \mathbf{v}_d$ of a $(d-1)$-simplex (and the associated spherical simplex). We shall denote the simplicial content, computed using either (5.116) or (5.119), using content$(\mathcal{V})$.

Let $\mathbf{v}_i$ and $\mathbf{v}_j$ denote the vectors in $\mathcal{V}$ that maximise dist$(\mathbf{v}_i, \mathbf{v}_j)$. It seems heuristically appropriate to assume that the longest edge makes one of the largest contributions to the defect between the planar and spherical contents. The mid-point of the edge from $\mathbf{v}_i$ to $\mathbf{v}_j$ is

$$\mathbf{v}_{\text{mid}} = \frac{\mathbf{v}_i + \mathbf{v}_j}{2}. \tag{5.120}$$

The projection of this point onto the surface of the sphere is associated with the vector

$$\mathbf{v}_* = \frac{\mathbf{v}_{\text{mid}}}{|\mathbf{v}_{\text{mid}}|}. \tag{5.121}$$

We shall propose that, if content$(\mathcal{V})$ is an approximation of the content of the hyperspherical simplex specified by the vertices in $\mathcal{V}$, then

$$\text{content}((\mathcal{V} \setminus \{\mathbf{v}_i\}) \cup \{\mathbf{v}_*\}) + \text{content}((\mathcal{V} \setminus \{\mathbf{v}_j\}) \cup \{\mathbf{v}_*\})$$

is consistently a better approximation.

5.4.6 Implementing a Subdivision Algorithm

The goal of this section is to add a subdivision block to the detector to approximate the conditional probabilities, given H_0 and H_1, of multiple interval events as they arrive. To approximate the orthant probability, the content of the $(d-1)$-spherical simplex—approximated by summing the content of many, smaller simplices—is divided by the (surface) content of the $(d-1)$-sphere that contains it. The content of the surface of the $(d-1)$-sphere is

$$C_{sph} = \frac{2\pi^{d/2}}{\Gamma(d/2)}, \tag{5.122}$$

where $\Gamma(\cdot)$ is the Gamma function.

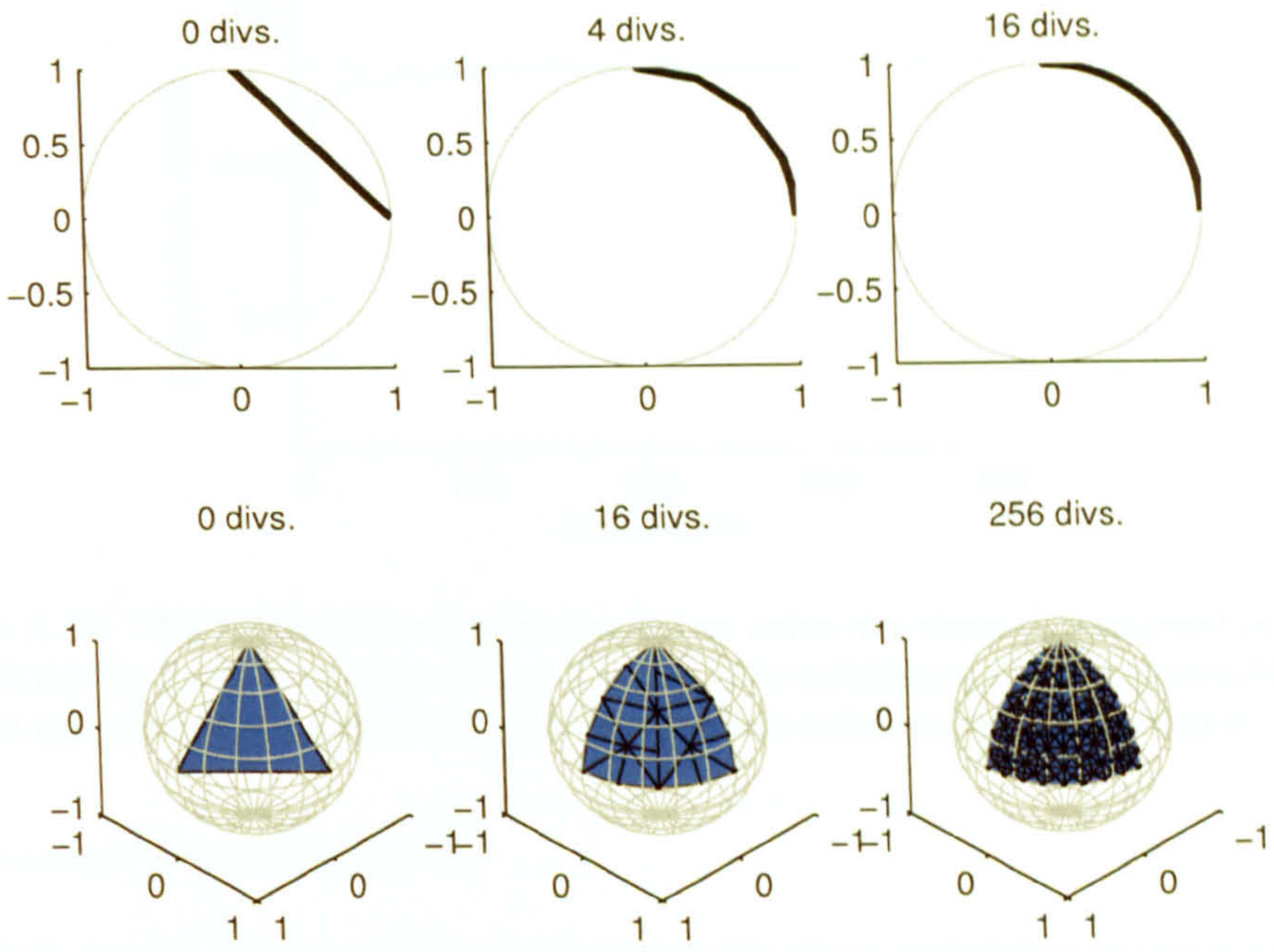


Figure 5.26: Top row: subdividing a line segment until it converges onto an arc segment. Bottom row: subdividing a plane triangle until it converges onto a spherical triangle.

The main difficulty in verifying that the subdivision approach works is that we cannot, in the general case, compare the approximation with any analytical results that have not themselves been approximated. However, at least four ways to proceed remain open: 1) plot the subdivision of the simplicies in two and three dimensions using MATLAB, to confirm that the algorithm at least carries out the expected geometric transformations; 2) compare the approximation with analytical results for a problem having a non-trivial covariance matrix and trivial dimensionality; 3) compare the approximation with analytical results for a problem having a non-trivial dimensionality and trivial covariance matrix; 4) compare the approximation with an empirical value.

Visualising Subdivision in Two and Three Dimensions

Figure 5.26 shows the subdivision of a line segment into progressively shorter line segments. After sixteen subdivisions, the chain of lines resembles a circular arc. Similarly, the subdivision of a plane triangle into progressively smaller triangles eventually resembles a spherical triangle. The algorithm behaves as expected and can now be applied to specific orthant probability and zero crossing problems.

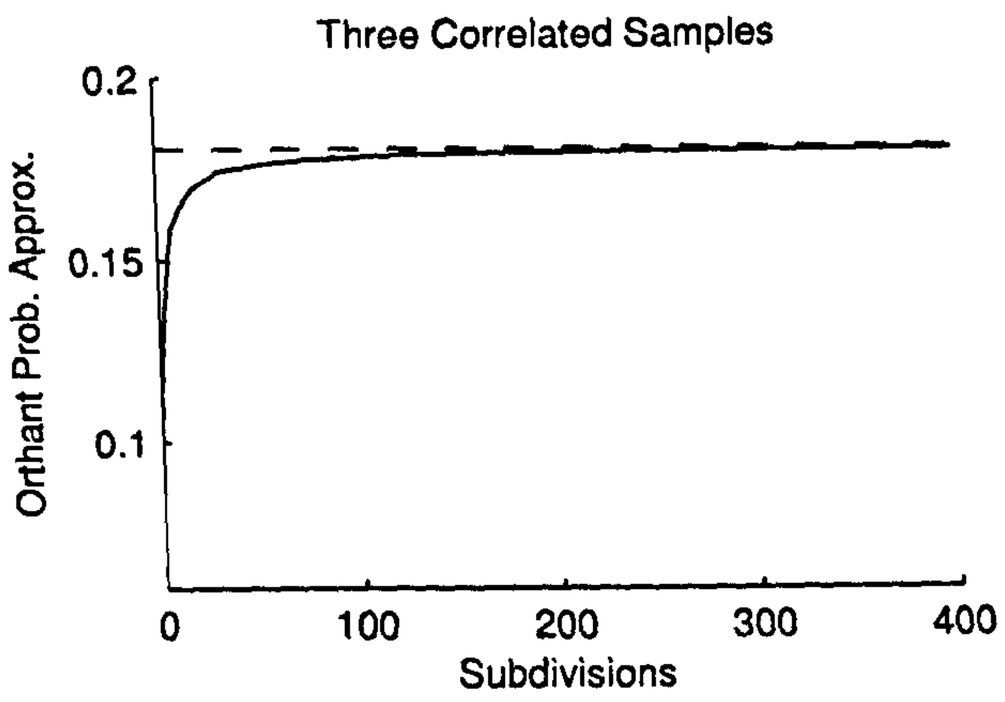


Figure 5.27: The subdivision algorithm used to solve the three-dimensional orthant probability for the covariance matrix in (5.123). The solid line plots the approximation against the number of subdivisions. The dashed line indicates the exact solution.

Three-sample Orthant Probability

We shall start by revisiting the three-dimensional orthant probability example worked out in Section 5.4.3. Let three samples of a Gaussian process, x_1, x_2 and x_3, have zero mean and covariance matrix

$$\Sigma_X = \begin{pmatrix} 1 & 0.9 & -0.22 \\ 0.9 & 1 & -0.2 \\ -0.22 & 0.2 & 1 \end{pmatrix}. \tag{5.123}$$

It can be verified by numerous methods that

$$P(x_1 \geq 0, x_2 \geq 0, x_3 \geq 0) = 2.2674.$$

Figure 5.27 illustrates how the approximation converges onto the true solution, as the simplex is subdivided four hundred times (i.e., there are 401 planar simplices at the end). The slow convergence to the solution is not encouraging.

Six-sample Orthant Probability

Let six samples of a Gaussian process, $x_1, \ldots, x_6$, have zero mean and zero covariance; in other words, the process is white. Because the samples are independent, we have the trivial result:

$$P(x_1 \geq 0, \ldots, x_6 \geq 0) = \frac{1}{2^6} = \frac{1}{64} \approx 0.0156.$$

Figure 5.28 illustrates how the approximated orthant probability converges *towards* the true solution, as the simplex is subdivided ten thousand times. However, even after such a large number of iterations, the solution is clearly inadequate.

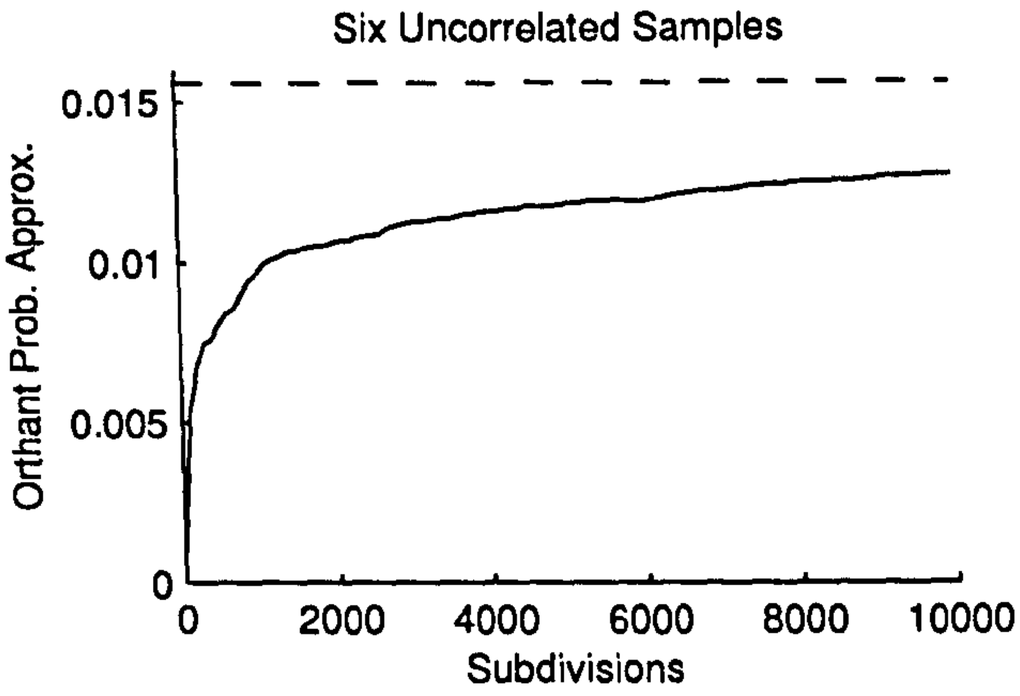


Figure 5.28: The subdivision algorithm used to solve the six-dimensional orthant probability for uncorrelated variables. The solid line plots the approximation against the number of subdivisions. The dashed line indicates the exact solution.

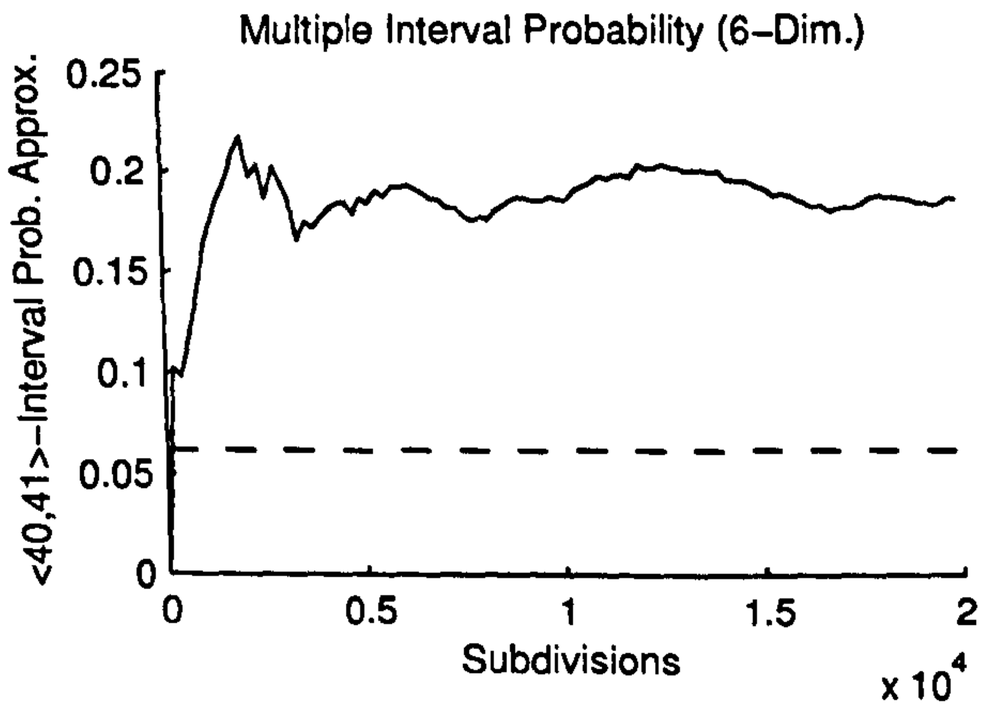


Figure 5.29: The subdivision algorithm used to solve the probability that, given a zero crossing has been observed, it is preceded by a 41-sample interval and a 40-sample interval. The solid line plots the approximation against the number of subdivisions. The dashed line indicates an empirical solution.

Joint Probability of Two Consecutive Intervals

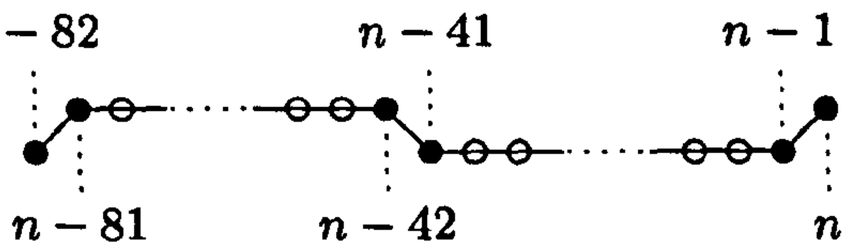
Last of all, we shall attempt to utilise the subdivision technique to approximate the probability of the multiple interval event $\langle i_1 = 40, i_2 = 41 \rangle$ in a Gaussian process described by the MGMM

$$\left\langle \frac{1}{2}, 40, 0, +2\pi \cdot 200, 0 \right\rangle + \left\langle \frac{1}{2}, 40, 0, -2\pi \cdot 200, 0 \right\rangle .$$

In other words, we wish to determine the probability that, given a zero crossing has just been received, it is immediately preceded by an interval 41 samples long, and a

second interval 40 samples long. (Note that we are dealing only with the noise-only hypothesis; there is no signal present.)

The detector responds to the following pattern of signed samples, or the same pattern with the signs reversed. (Empty circles effectively represent samples that are ignored.)



Labelling the six filled samples from right-to-left, i.e., $x_1 = x[n], \ldots, x_6 = x[n-82]$, the correlation matrix governing the samples is

$$\Sigma_X = \begin{pmatrix} 1 & \rho_X[1] & \rho_X[41] & \rho_X[42] & \rho_X[81] & \rho_X[82] \\ \rho_X[1] & 1 & \rho_X[40] & \rho_X[41] & \rho_X[80] & \rho_X[81] \\ \rho_X[41] & \rho_X[40] & 1 & \rho_X[1] & \rho_X[40] & \rho_X[41] \\ \rho_X[42] & \rho_X[41] & \rho_X[1] & 1 & \rho_X[39] & \rho_X[40] \\ \rho_X[81] & \rho_X[80] & \rho_X[40] & \rho_X[39] & 1 & \rho_X[1] \\ \rho_X[82] & \rho_X[81] & \rho_X[41] & \rho_X[40] & \rho_X[1] & 1 \end{pmatrix}, \tag{5.124}$$

where $\rho_X[\cdot]$ is the sampled autocorrelation function for the process. The probability that all the filled samples are *positive* is the all-positive orthant probability computed for Σ_X. However, we should like to determine the probability that the samples form the pattern $(-,+,+,-,-,+)$ or $(+,-,-,+,+,-)$, as opposed to $(+,+,+,+,+,+)$. This is equivalent to twice the all-positive orthant probability for the correlation matrix

$$\Sigma_{\text{neg}} \Sigma_X \Sigma_{\text{neg}}^T,$$

where Σ_{neg} is the diagonal matrix

$$\Sigma_{\text{neg}} = \begin{pmatrix} 1 & & & & & \\ & -1 & & & & \\ & & -1 & & & \\ & & & 1 & & \\ & & & & 1 & \\ & & & & & -1 \end{pmatrix}. \tag{5.125}$$

Figure 5.29 shows how the approximation evolves as the simplices are subdivided twenty thousand times. Two surprising features are present in this graph. First, the approximation *exceeds* the true solution throughout most of the run. In two and three dimensions, this is impossible, and it seems reasonable to conjecture that the same ought to be true for all dimensionalities. Second, in many iterations, the subdivision of a simplex leads the approximation to grow *smaller*; this, too, is impossible in two and three dimensions.

The condition numbers of the matrices used to evaluate the Cayley-Menger determinant upon examination are, in most cases, far to high to permit accurate computation of the determinant. In summary, the inaccuracies that arise during the process of evaluating simplicial contents accumulate to the point where this kind of approach becomes highly impractical on a standard desktop computer. Besides, even if the approximation were to converge succesfully, the computational cost of evaluating the probability of a short sequence of intervals every time a zero crossing arrives rules out most likely applications in sonar.

Attempts to evaluate the probability of many zero crossing intervals in a short time period is hindered by the intractable relationship between i) the correlation amongst the amplitude samples of a process and ii) the correlation amongst its zero crossing intervals. As we have seen, this relationship can be fully described with reference to the analogues of triangles and spheres in high dimensions, but an analytical method to discover the content of these shapes is lacking. In the final section, we shall take the opposite approach: measure fewer, independent intervals over a longer time period.

5.5 Post-detection Integration

The first four sections of this chapter each independently tackled a particular problem or restriction facing the detection routines carried over from Chapter 4. In this fifth and final section, we shall attempt to combine the best aspects of these modifications into a single receiver, i.e., one which i) fixes the false alarm rate; ii) detects sinusoids; iii) incorporates information from the envelope and fine structure and iv) incorporates information from many intervals.

The key idea that will allow us to draw these four strands together is *post-detection integration*: the averaging of many independent, identically-distributed samples to obtain Gaussian statistics by the central limit theorem (Peebles, 1993). Averaging samples until they converge upon a normal distribution is a standard aspect of conventional sonar detection and permits the construction of detectors capable of operating at arbitrarily low SNRs. (This topic was briefly discussed in Chapter 3.) Our failure in the preceding section to find an expression for the joint p.d.f. governing a sequence of zero crossing intervals on short time scales adds to the motivation for an averaging approach.

The prototypical narrowband power detector consists of four components: a linear analysis filter, a square-law device (a block whose output is the squared-envelope of its input), a linear post-detection filter, and a threshold (Burdic, 1984). We shall augment this model by adding a pathway that measures the zero crossing intervals of the analysis filter output and supplies them to a second post-detection filter. The likelihood functions that make up the decision rule will then be bivariate Gaussian probability density functions. These considerations give rise to the kind of assembly schematically illustrated in Figure 5.30.

The post-detection filters are designed to average the input signal to produce an output signal whose samples are Gaussian-distributed; however, the input samples—whether from a square-law device or a zero crossing interval device—are generally correlated on short time scales. The means and variances of the Gaussian likelihood functions must be known before they can be placed into a decision rule. In theoretical terms, the variance of the samples at the post-detection filter output is minimised when the samples are *statistically independent*. The simplest means of guaranteeing independence between samples is to insert suitably long waiting periods between each measurement.

We shall examine the components of the algorithm depicted in Figure 5.30 in the following order: the upper branch, which extracts and averages the squared envelope; the lower branch, which extracts and averages zero crossing intervals; and lastly, the decision rule block, which combines both envelope and timing information to decide whether or not a signal is present. Once the detector has been fully developed, a series of experiments can be undertaken to assess its performance. The results will be presented as ROC curves.

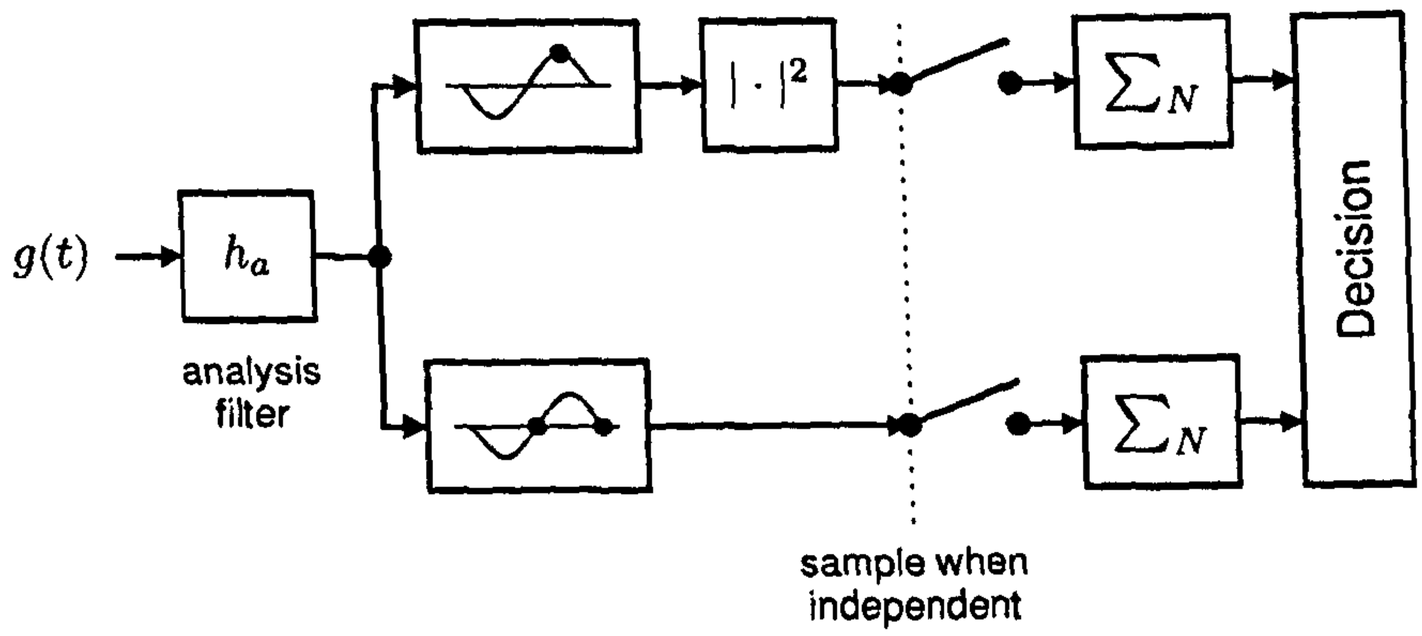


Figure 5.30: Joint interval detector with post-detection averaging. The upper branch averages N independent samples of the squared-envelope. The lower branch averages N independent zero crossing intervals.

5.5.1 Squared-envelope Detection Branch

Noise Only

If the input to the analysis filter is noise with autocovariance function $\gamma_X(\tau)$, then the probability density function governing the squared-envelope output, prior to averaging, is that of the exponential distribution.

$$p_E(e) = \frac{1}{2\gamma_X(0)} \exp\left(\frac{-e}{2\gamma_X(0)}\right). \tag{5.126}$$

The mean and variance of this distribution are as follows (Peebles, 1993).

$$\text{mean}\{E \mid H_0\} = 2\gamma_X(0) \tag{5.127}$$
$$\text{var}\{E \mid H_0\} = 4\gamma_X^2(0). \tag{5.128}$$

The p.d.f. is shown as a solid line in the top-left plot of Figure 5.31.

The next block averages N samples. When N is large, the averaged statistic starts to converge towards a Gaussian distribution with mean $2\gamma_X(0)$ and variance $4\gamma_X^2(0)/N$. The top row of plots in Figure 5.31 show how the probability density function tends towards a Gaussian shape as N increases.

Using E_a to denote the averaged sample, when $N \gg 1$,

$$p_{E_a}(e_a) \approx \frac{1}{\sqrt{2\pi \cdot 4\gamma_X^2(0)/N}} \exp\left(\frac{-(e - 2\gamma_X(0))^2}{8\gamma_X^2(0)/N}\right). \tag{5.129}$$

Signal and Noise

When a sinusoid with amplitude A is added to the input signal, the samples at the output of the analysis filter are non-central chi-squared-distributed (Whalen, 1971, see

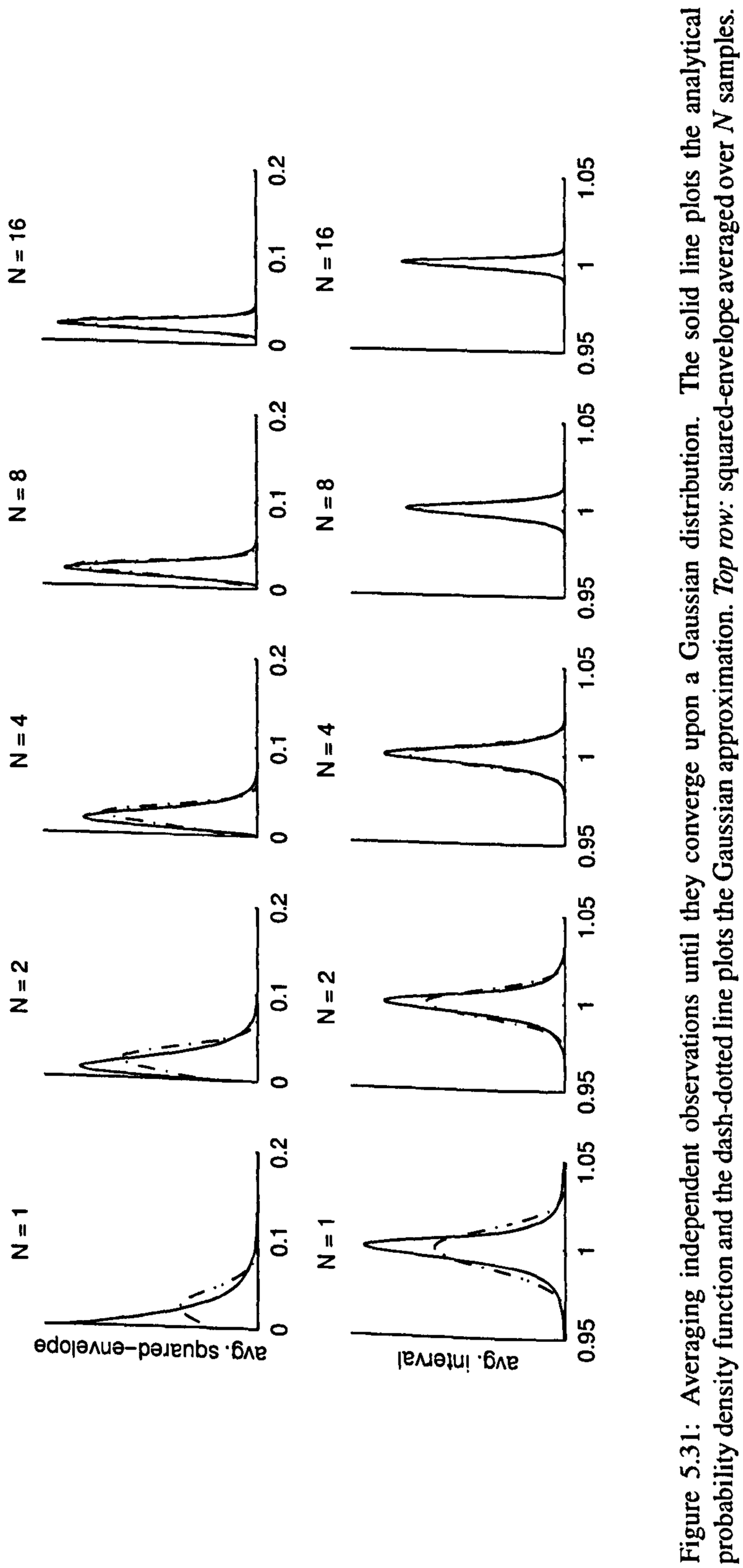


Figure 5.31: Averaging independent observations until they converge upon a Gaussian distribution. The solid line plots the analytical probability density function and the dash-dotted line plots the Gaussian approximation. *Top row:* squared-envelope averaged over N samples. *Bottom row:* continuous intervals averaged over N samples.

also §1.1.1). In this case, the probability density function governing E is

$$p_E(e) = \frac{1}{2\gamma_X^2(0)} \exp\left(\frac{e + |\mathscr{H}_a(\omega_c)|^2 A^2}{-2\gamma_X^2(0)}\right) I_0\left(\frac{|\mathscr{H}_a(\omega_c)| A\sqrt{e}}{\gamma_X^2(0)}\right), \tag{5.130}$$

where $|\mathscr{H}_a(\omega_c)|$ is the linear magnitude response of the analysis filter at the sinusoid frequency. The mean and variance of this distribution are as follows.

$$\text{mean}\{E \mid H_1\} = 2\gamma_X(0) + A^2 \tag{5.131}$$

$$\text{var}\{E \mid H_1\} = 4\gamma_X(0)[\gamma_X(0) + A^2]. \tag{5.132}$$

Naturally, (5.131) and (5.132) reduce to (5.127) and (5.128), respectively, when the signal amplitude is zero. The test statistic E_a, produced by averaging N independent samples, has mean $2\gamma_X(0)+A^2$ and variance $4\gamma_X(0)[\gamma_X(0)+A^2]/N$, and approaches a Gaussian distribution when $N \gg 1$.

5.5.2 Continuous Interval Detection Branch

We now turn our attention to the lower branch of Figure 5.30, which extracts $N + 1$ consecutive zero crossing times using linear interpolation and computes the average of the N resulting intervals. This random variable is designated I_{ca} (i.e., I_c, averaged).

Noise Only

The probability density and cumulative distribution functions governing a single zero crossing interval of a continuous-time, wide-sense stationary Gaussian noise process were approximated in Section 4.4.3, and the former was found to be

$$p_{I_c}(i_c) = \begin{cases} \dfrac{(\rho^2(i_c) - 1)\rho''(i_c) - \rho(i_c)(\rho'(i_c))^2}{2(\rho^2(i_c) - 1)^{3/2}\sqrt{\rho''(0)}} & \tau_0 < i_c < 2\tau_0 \\ 0 & \text{otherwise.} \end{cases} \tag{5.133}$$

Allowing for the moment that (5.133) is precisely correct, the mean and variance of the distribution are as follows.

$$\text{mean}\{I_c \mid H_0\} = \int_{\tau_0}^{2\tau_0} i_c p_{I_c}(i_c) di_c \tag{5.134}$$

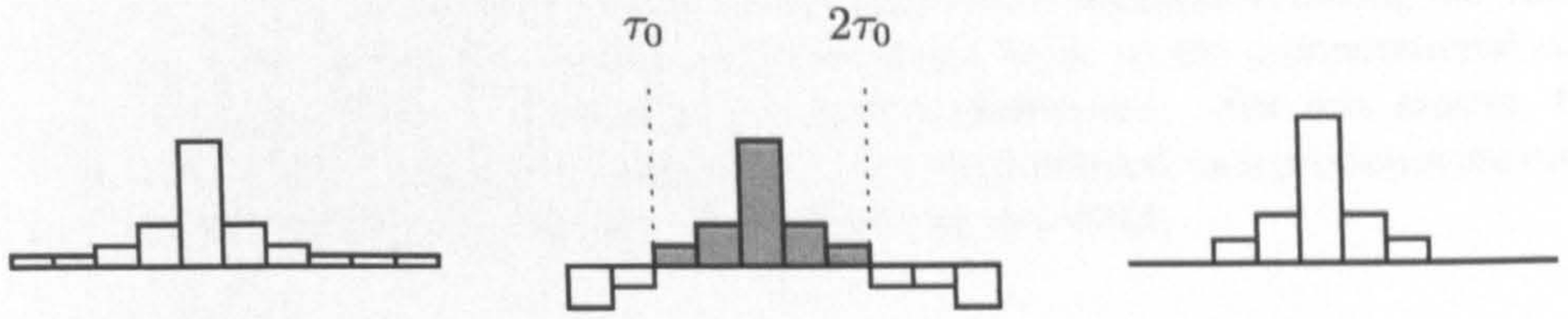
$$\text{var}\{I_c \mid H_1\} = \int_{\tau_0}^{2\tau_0} i_c^2 p_{I_c}(i_c) di_c - \left(\int_{\tau_0}^{2\tau_0} i_c p_{I_c}(i_c) di_c\right)^2. \tag{5.135}$$

Neither of these integrals has an immediately evident analytical solution[1], but a suitable approximation can be obtained by numerical means. The mean and variance of the

[1] According to Rice's formula (Rice, 1944), the average number of zero crossings per unit time for a wide-sense stationary Gaussian process is $\frac{1}{\pi}\sqrt{-\rho''(0)}$. Consequently, the mean zero crossing interval is exactly

$$\text{mean}\{I_c \mid H_0\} = \frac{\pi}{\sqrt{-\rho''(0)}}.$$

There is no similar formula for the variance.



A. True p.d.f. B. Approximate p.d.f. C. Adjusted p.d.f.

Figure 5.32: Mitigating the effect of interval aliasing on estimates of the mean and variance. A) illustration of the true distribution; B) the tails of the analytical distribution are inaccurate, but the majority of its support is accurate (grey blocks); C) assume that intervals are only received in the valid region, and normalise this region to unit area.

distributions are found by evaluating the integrals as summations, using a sufficiently fine-grained integration step. Before a suitable approximation of the mean and variance can be computed, the probability density function in (5.133) must be normalised to unit area, to overcome the adverse effects of interval aliasing. (See Figure 5.32 and caption.)

Once the mean and variance of the interval distribution have been determined, they can be used to parameterise a Gaussian distribution. For a single sample (i.e., $N = 1$), there is a discrepancy between the Gaussian density function and the interval density function; most notably, the latter is more peaked than the former. The lower row of plots in Figure 5.31 show that, as N increases, the distribution of I_{ca} begins to approach Gaussian.

Signal and Noise

The signal-and-noise hypothesis, H_1, assumes that a sinusoid of amplitude A has been received in addition to noise. Two practical techniques for determining the interval distribution for a sine-in-noise mixture were discussed in Section 5.2. First, the sinusoid could be treated as though it were an ultra-narrowband Gaussian process. This approach inevitably introduces modelling errors, as the samples of a sinusoid do not follow a Gaussian distribution. However, the results of simple detection experiments in Section 5.2.7 suggest that the approach is acceptable if the signal-to-noise ratio is suitably low (< 20 dB). And second, a "Rayleigh-sum approach", the details of which are described in Section 5.2.5, can be used to approximate the probability density function with greater accuracy. This method of finding the interval distribution does not require the assumption of a particularly high or low SNR, but it is more costly in computational terms.

It is left to the designer of the detector to decide which of the two options above will be used to generate the signal-and-noise distribution. The first option is admissible for low or high signal-to-noise ratio applications, respectively, and requires less computational effort than the second option. However, the sole purpose of computing the signal-and-noise distribution is to obtain its first two moments, i.e., the mean and variance, which are then used to construct Gaussian likelihood functions. Both the computation

of the p.d.f. and its subsequent numerical integration are undertaken during the initial calibration phase, rather than during the operational loop, so the computational cost of producing the density function is a minor consideration. For this reason, the second option—the Rayleigh-sum approach—is to be preferred, as it produces the most reliable probability density functions and does so for any SNR.

5.5.3 Decision Rule

The decision rule operates on the average squared-envelope (peak squared-amplitude) and zero crossing interval received from the upper and lower pathways in Figure 5.30. As we have noted, when N is sufficiently large, E_a and I_{ca} converge to univariate Gaussian distributions. The central limit theorem also states that the joint statistic, $\langle E_a, I_{ca} \rangle$, will converge to a bivariate Gaussian distribution. For the hypothesis H_j, the bivariate Gaussian distribution, $p_{I_{ca}E_a}(i_{ca}, e_a \mid H_j)$, is completely described by a 2-element mean vector, $\vec{\mu}_j$ and a 2×2 covariance matrix, Σ_j. The decision rule of the detector is a hypothesis test that chooses H_1 if

$$\frac{p_{I_{ca}E_a}(i_{ca}, e_a \mid H_1)}{p_{I_{ca}E_a}(i_{ca}, e_a \mid H_0)} > \eta, \tag{5.136}$$

or, after placing Gaussian p.d.f.s in (5.136) and taking logs,

$$(\mathbf{x} - \vec{\mu}_0)^T \Sigma_0^{-1} (\mathbf{x} - \vec{\mu}_0) - (\mathbf{x} - \vec{\mu}_1)^T \Sigma_1^{-1} (\mathbf{x} - \vec{\mu}_1) > \ln \frac{|\Sigma_1|}{|\Sigma_0|} + 2 \ln \eta, \tag{5.137}$$

and H_0 otherwise.

The conditional means and variances of the distribution are computed according to Sections 5.5.1 and 5.5.2 above. The off-diagonal elements of the covariance matrix are the covariance of a zero crossing interval and its peak squared-amplitude. In the experiments that follow, the covariances are set to zero, so that peaks and intervals are modelled as statistically independent. Although this simplification could decrease the performance of the detector, there are at least two reasons to suppose that a diagonal covariance matrix is acceptable. First, empirical measurements of the covariance between intervals and peaks show that their correlation coefficient[1] is quite low under most circumstances (≈ 0.01). Second, it is difficult to estimate the covariance between intervals and peaks to a suitable degree of accuracy, when the estimates are computed numerically from the joint probability density function, which has in the first instance been obtained using a series of approximations (*cf.* Section 5.2).

[1] It could be objected that modelling the statistical dependency between intervals and peaks significantly improved detection performance under some conditions (e.g., see Figure 5.18), and that by neglecting the covariance term, this benefit will be lost. Against this, it must be remembered that although the intervals and peaks exhibit a strong *statistical dependence*, they do not share a strong *linear relationship*. Considering an alternative example, X and Y are statistically dependent if

$$p_{XY}(x, y) = \begin{cases} 1/(10\pi) & x \geq 0 \text{ and } 4 \geq \sqrt{x^2 + y^2} \geq 6 \\ 0 & \text{otherwise.} \end{cases}$$

Nevertheless, X and Y are uncorrelated, i.e., $E\{XY\} = 0$.

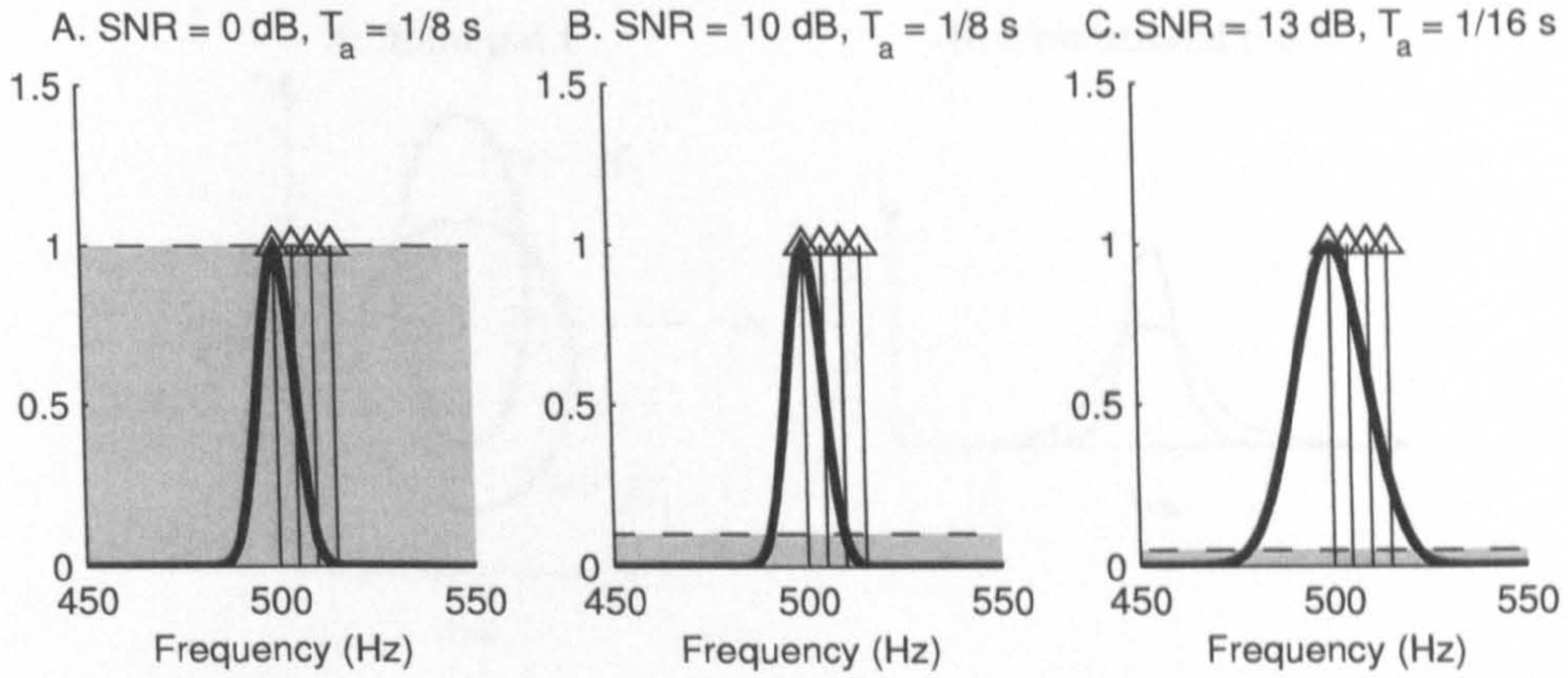


Figure 5.33: Signal and noise conditions. A) On-centre and off-centre detection at 0 dB narrowband SNR; for results, see Figs. 5.36 and 5.37. B) Identical to (A), but with SNR raised to 10 dB; for results, see Fig. 5.38. C) Identical to (B), but with bandwidth doubled and pre-analysis SNR adjusted to match post-analysis SNR; for results, see Fig. 5.39.

5.5.4 Experimental Results and Analysis

The performance of the joint interval-peak detector with post-detection integration was assessed by a signal detection task. Two signals were synthesised, the first consisting of white noise with no signal (H_0), and the second a mixture of a sinusoid and noise (H_1), and the detector was configured to choose H_0 or H_1 according to a fixed likelihood threshold. The empirical true positive and false positive probabilities sampled from 400,000 trials[1] were used to generate each ROC curve. In order to compare the performance of the joint interval detector with that of a purely power-based approach, the ROC curves for peak squared-amplitude detector were also plotted. (This detector effectively omits the lower branch of the diagram in Figure 5.30.)

On-centre Detection

The first detection task required the detection of a sinusoid centred on an analysis band with an impulse response described by the MGMM

$$\left\langle \frac{1}{2}, 20, 0, +2\pi \cdot 500, 0 \right\rangle + \left\langle \frac{1}{2}, 20, 0, -2\pi \cdot 500, 0 \right\rangle . \tag{5.138}$$

The squared-magnitude response of this filter, normalised to unit peak gain, is plotted in Figure 5.33A, along with the (pre-filter) signal and noise power spectral densities. The integration stage averages $N = 64$ independent samples (i.e., interval-peak pairs) before applying the decision rule. The ROC curves that result are plotted in Figure 5.36. No consistent difference in performance between the joint interval-peak detector and

[1] Equivalent to about seven hours of sound, or approximately 25 million zero crossing intervals.

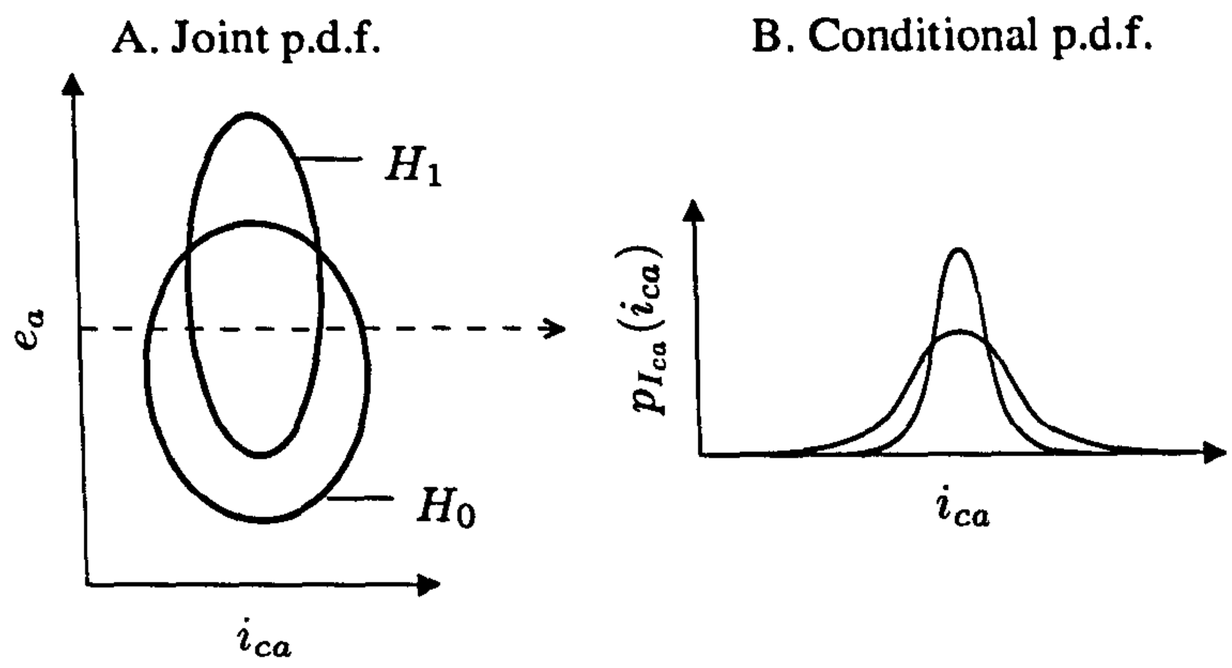


Figure 5.34: The decision based on $\langle e_a, i_{ca} \rangle$ can be envisaged as two separate stages. A) Illustrative contours of the joint conditional p.d.f.s for H_0 and H_1; B) once e_a has been measured, the interval p.d.f. conditioned on $E = e_a$ ought to add to enhance the decision. This is tantamount to choosing j to maximise $P(e_a \mid H_j)P(i_{ca} \mid e_a, H_j)$.

power detector is apparent, which is unsurprising in light of the results from earlier experiments that considered detection of a signal on the band centre. We also observe that by combining multiple samples in the detector, it is possible to secure lower false alarm probabilities than for an individual interval.

The second detection task held the narrowband signal-to-noise ratio fixed at 0 dB and varied the number of independent samples combined prior to the decision rule (N). The ROC curves that result are plotted in Figure 5.37. Once again, the signal is centred on the analysis band, so there is no significant processing gain afforded by incorporating temporal information. For moderate false alarm rates (e.g., 10^{-2}), it is possible to gain large increases in detection probability by increasing N, at the expense of longer integration periods.

There is one anomolous result concerning the on-centre (500 Hz) signal condition that must be accounted for, before proceeding to the final experiment. We know that the variance of the zero crossing intervals is reduced when a signal is added at the centre of the analysis band. This can be confirmed analytically and empirically. Furthermore, because the intervals and peaks are (almost) uncorrelated, the difference in interval variance between H_0 and H_1 must, in theoretical terms, assist the detector to some degree, as the diagram in Figure 5.34 illustrates. Yet this improvement is not visible in Figure 5.38.

This apparent contradiction is resolved if the improvement in detection performance is *negligible*—and hence literally invisible in the figure—as opposed to non-existent. It can be shown that this is in fact the case for on-centre detection with post-detection averaging. The conditional means for the on-centre condition may be treated as equal[1], with the counterintuitive consequence that averaging many intervals fails to increase

[1] Adding a signal to the band centre causes the mean interval to change by an exceedingly small amount. For the on-centre condition shown in Figure 5.33A, the difference is approximately 3 μs.

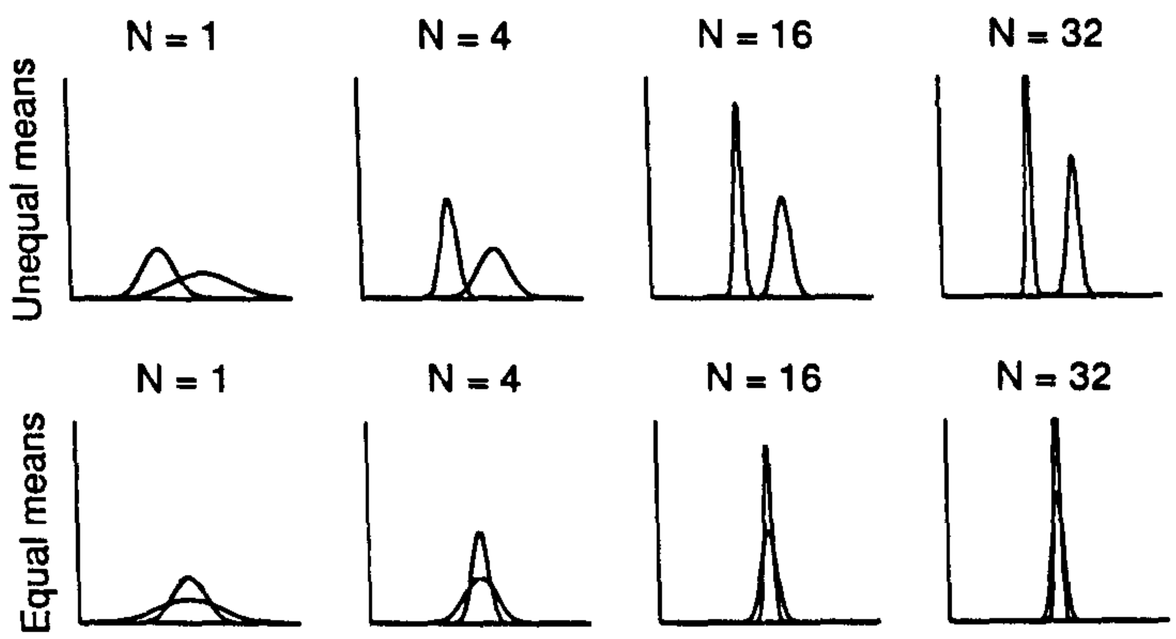


Figure 5.35: A) If the means of two conditional distributions are unequal, then averaging N samples improves detection. The probability density functions grow narrower and more Gaussian-like, but their means remain separated by the same amount; thereby, the detector can more reliably separate incoming samples into two classes. B) If the means of the distributions are equal, then averaging N samples merely causes both distributions to shrink onto the same centre; separability is unaffected.

detection performance to any extent. On the other hand, the conditional means of the envelope test statistic, e, are unequal; thus, increasing N monotonically improves detection performance. The graphs in Figure 5.35 illustrate how a separation in conditional means is required, if averaging multiple samples is to improve the quality of the detector. Placing the signal away from the band centre causes the conditional mean intervals to diverge. The final experiments investigate the detection of a signal that has been displaced from the band centre.

Off-centre Detection

In the third detection task, the signal-to-noise ratio was held at 0 dB, and 64 samples were averaged. The performance of the detectors was recorded for various signal displacements, specifically, 0 Hz (no displacement), +5 Hz, +10 Hz and +15 Hz. The power spectral densities for these four signals are shown in Figure 5.33A. The ROC curves that were produced are shown in Figure 5.38, with each condition plotted on a separate set of axes. There is no appreciable difference in performance between the two detectors when the signal is located at 500 Hz or 515 Hz, but the joint interval-peak detector achieves a small increase in performance over the power detector when the signal frequency is 505 Hz or 510 Hz. Evidently, at 500 Hz, the performance increase contributed by temporal information is negligible—as explained above—and at 515 Hz, signal attenuation reduces the detector to near-chance performance.

For signal and noise mixed at 0 dB narrowband SNR, only a marginal performance increase follows from incorporating temporal statistics into the detection process. In order to accentuate this benefit, either the number of samples, N, could be increased, or the signal-to-noise ratio could be raised. The fourth detection task pursues the

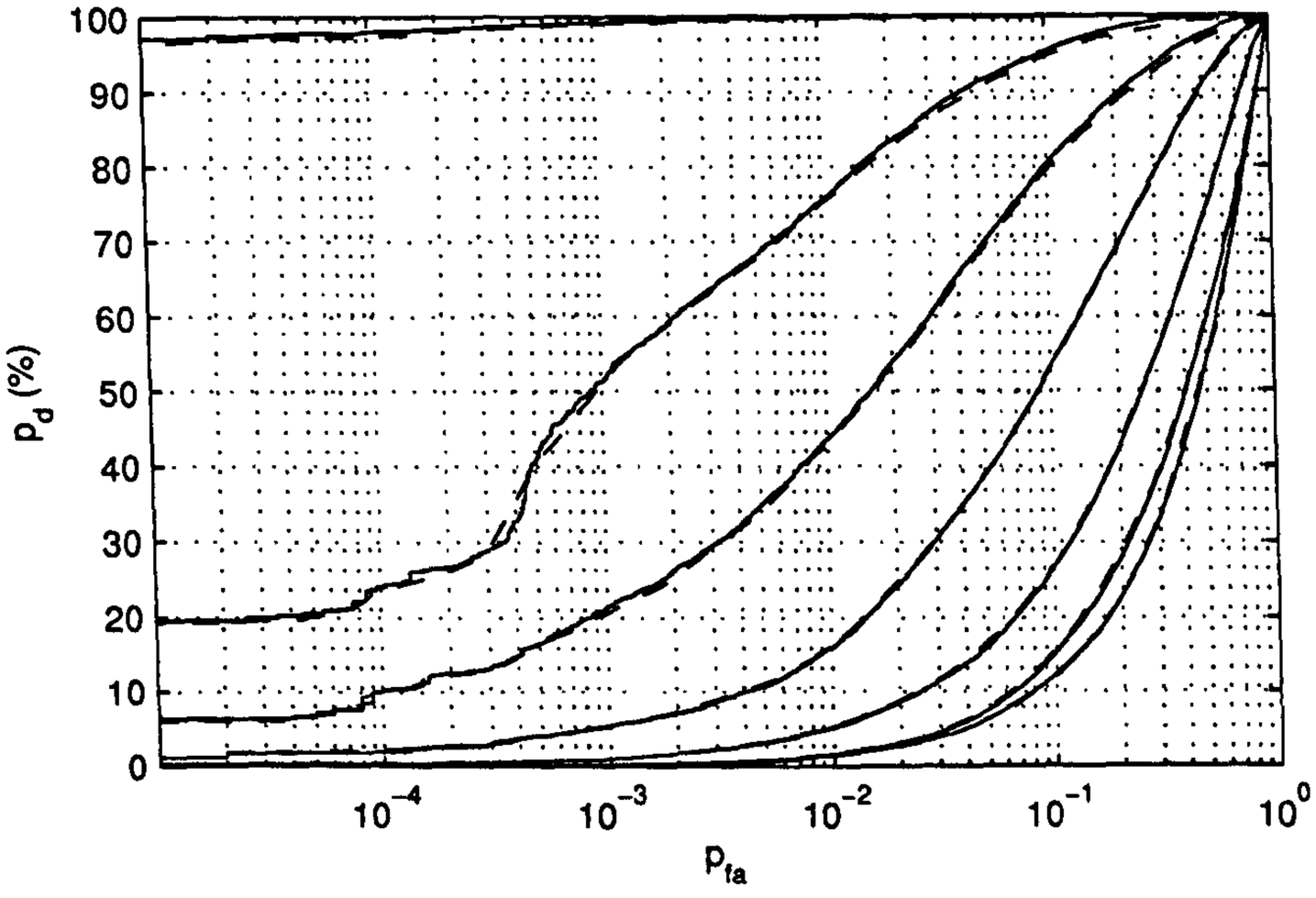


Figure 5.36: ROC curves for on-centre detection with 64 independent samples. The curves shown are for narrowband SNRs -10 dB (lowest curve), -5 dB, 0 dB, 3 dB, 5 dB, 7 dB, and 10 dB (uppermost curve), where a solid line indicates the performance of the power ("peak-only") detector and a dashed line indicates the performance of the joint interval-peak detector.

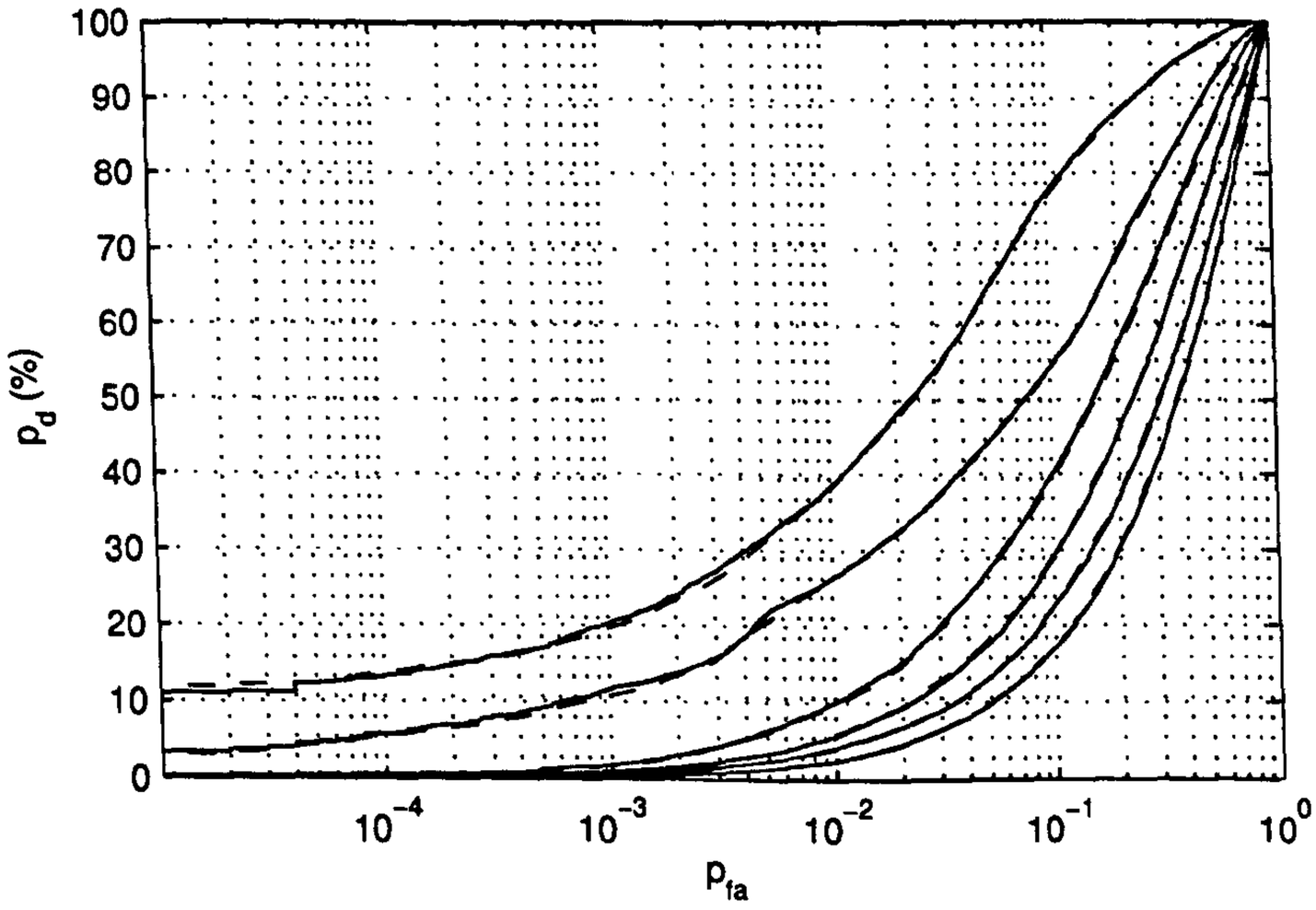


Figure 5.37: ROC curves for on-centre detection with 0 dB narrowband SNR. The curves shown are for 16 (lowest curve), 32, 64, 128, 256 and 512 (uppermost curve) samples, respectively. See Figure 5.36 caption above for key.

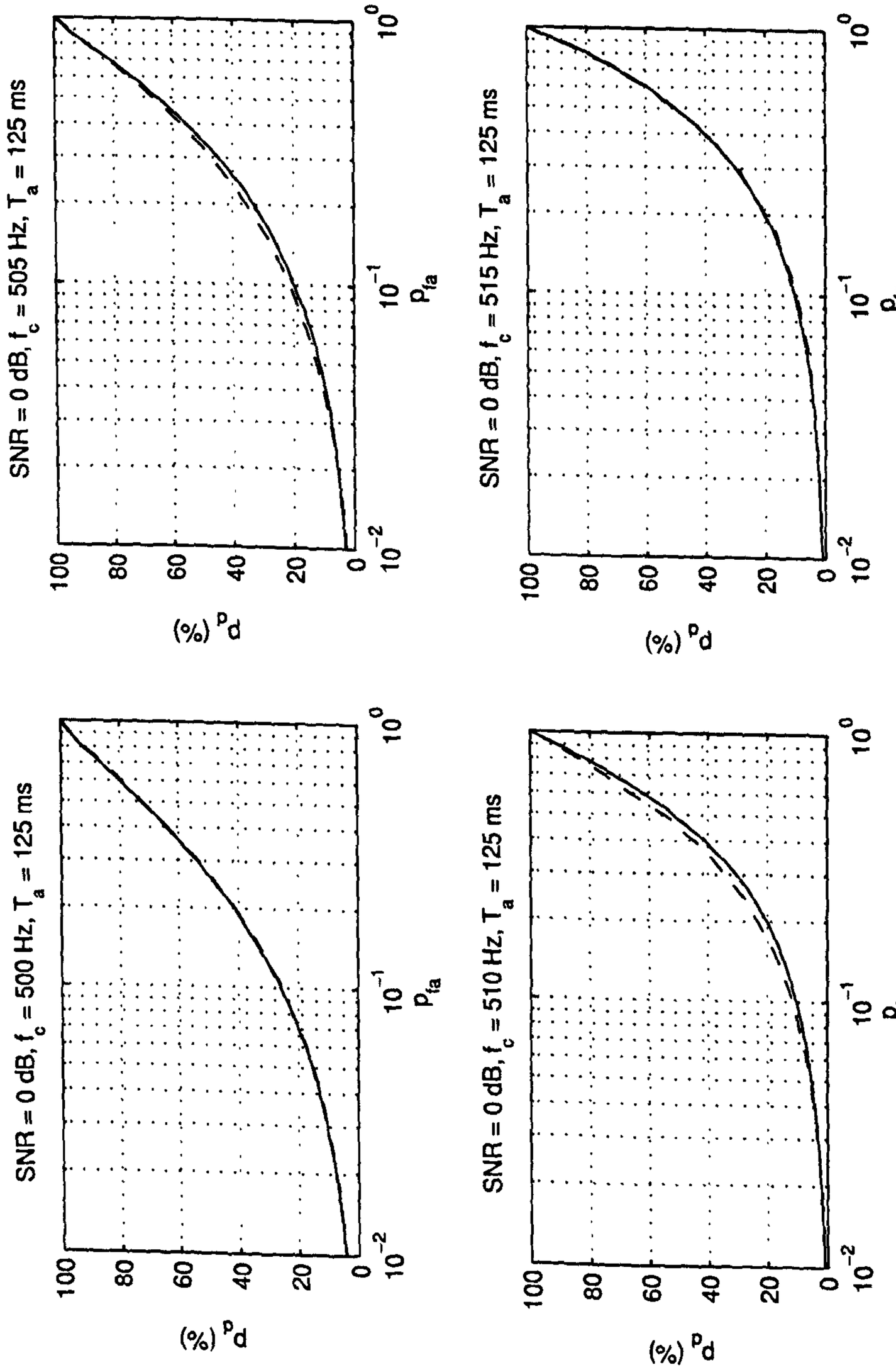


Figure 5.38: ROC curves for off-centre detection at 0 dB narrowband signal-to-noise ratio.

second of these options. Figure 5.38 plots the performance of the joint interval-peak and power detector when the SNR is raised to 10 dB. The signal and noise conditions are plotted in Figure 5.33B. Supplying evidence from zero crossing intervals to the decision rule leads to a pronounced increase in detection probability when the signal is placed at 505 Hz or 510 Hz. In the 505 Hz signal frequency condition, at low false alarm probabilities ($\lesssim 0.0005$), the joint peak-interval detector is inferior to the power detector (Figure 5.39, top-right). The failure of the detector in this instance can be attributed to a poor approximation of the tail of the interval distribution. Some model error is unavoidable: the Gaussian distribution has infinite support, which means that some probability must be assigned to negative zero crossing intervals.

The fifth and final detection task required the detection of a signal in a wider analysis bandwidth. The signal-to-noise ratio was adjusted so that the post-analysis noise power was matched with the preceding experiment, which equates to a pre-analysis SNR of approximately 13 dB for the on-centre condition. The signal and noise conditions are shown in Figure 5.33C, and the corresponding ROC curves are plotted in Figure 5.40. It is seen that, except for the model errors introduced by fixing the false alarm probability too low, off-centre detection improves considerably when envelope and zero crossing interval measurements are combined in the detector.

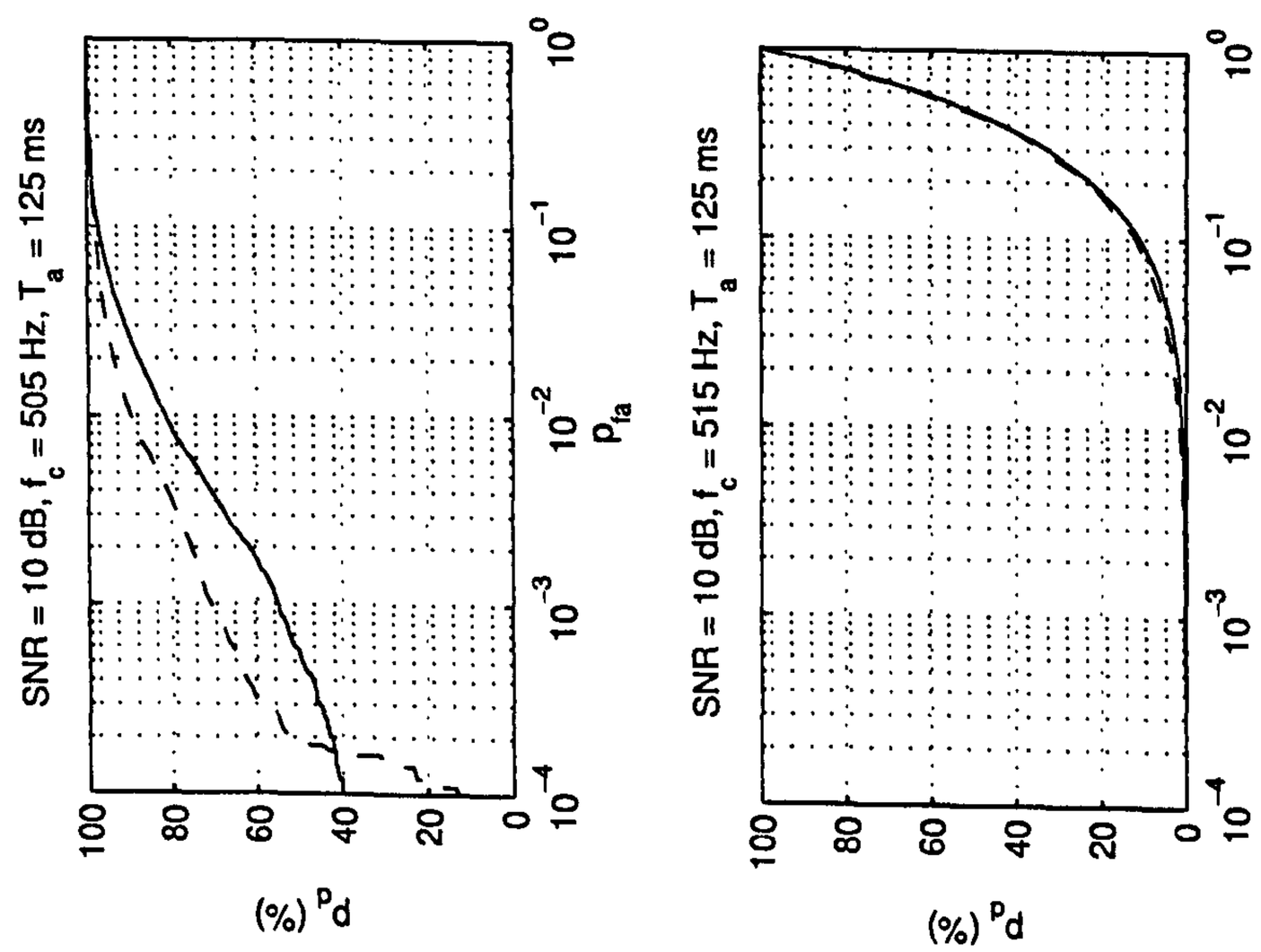


Figure 5.39: ROC curves for off-centre detection at 10 dB narrowband signal-to-noise ratio.

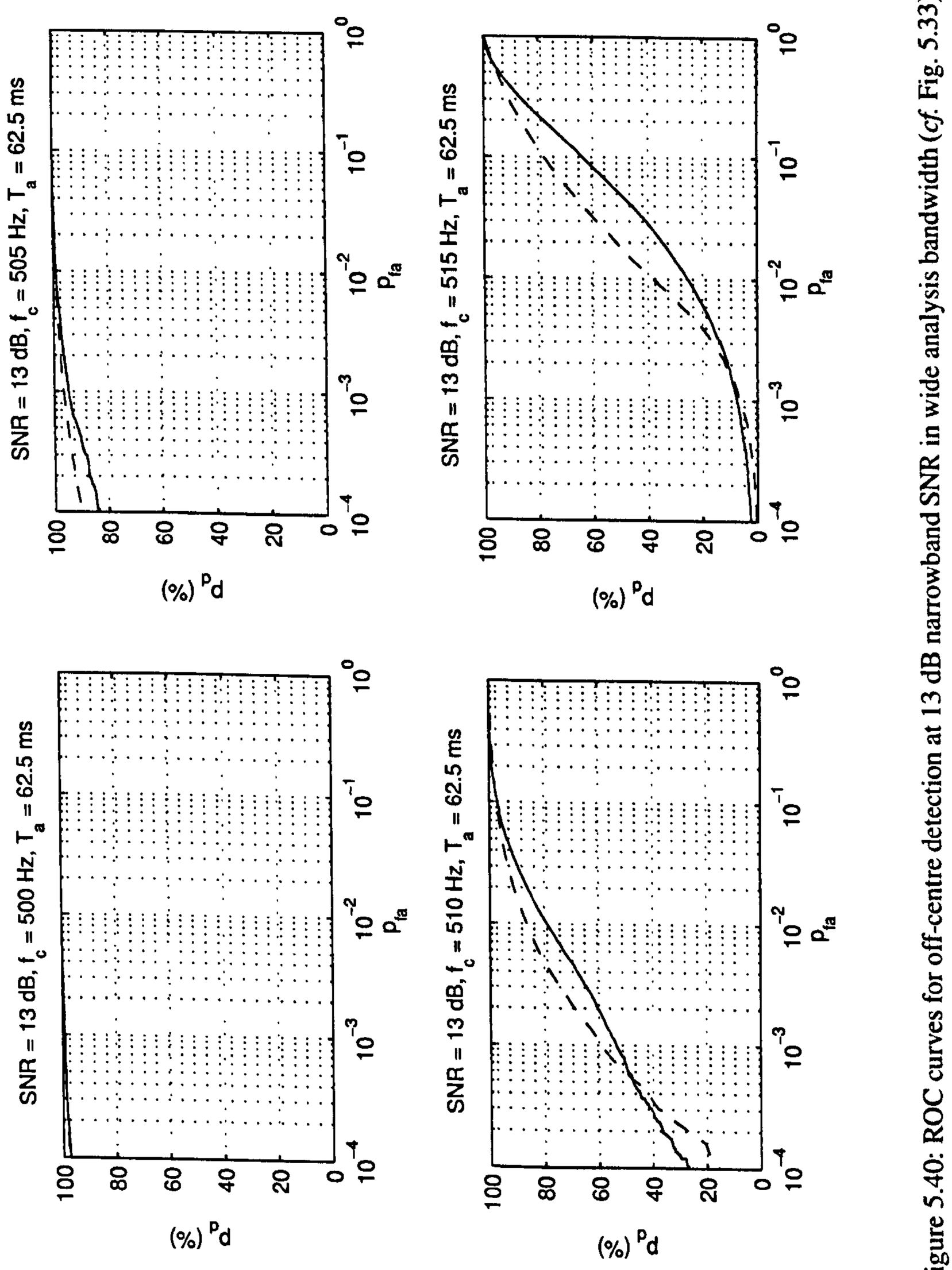


Figure 5.40: ROC curves for off-centre detection at 13 dB narrowband SNR in wide analysis bandwidth (*cf.* Fig. 5.33).

5.6 Summary

The aim of this chapter was to bridge the conceptual gap between the capabilities of the elementary interval detectors in Chapter 4, and the envisaged practical requirements of ZCPA-based applications in Chapter 6. Four assumptions from the previous chapter have been reappraised.

The first assumption concerned the metric used to evaluate the performance of a detector. Chapter 4 considered only minimum error detectors, namely, receivers which assign an equal cost to false alarms and false dismissals. This chapter investigated interval detectors that fulfil the Neyman-Pearson criterion, that is, to maximise the detection probability for a constant false alarm rate. Whereas a power detector applies a single threshold to its test statistic, an interval detector requires two: one to reject intervals which are too short, the other to reject intervals which are too long.

The second assumption related to the kind of random process that the detector could choose between. Chapter 4 required that each hypothesis be a wide sense stationary Gaussian process, and the derivation of the interval distribution incorporated this assumption. The most important random process discussed in the sonar literature, after pure Gaussian noise, is a randomly-phased sinusoid in Gaussian noise. This process has non-Gaussian samples, and consequently, the distribution of its zero crossing intervals could not be found using the techniques in Chapter 4.

In this chapter, we exploited the fact that a closely-related random process, namely, a randomly-phased sinusoid with a Rayleigh-distributed amplitude, was both stationary and Gaussian. By choosing a suitable linear combination of the interval distributions for this type of process, the distribution for a sinusoid with a constant amplitude was approximated. The interval detector based on this likelihood function was able to optimally detect a sinusoid in noise and outperform a squared-envelope detector when the signal was sufficiently displaced from the band centre.

The third restriction in Chapter 4 stated that an interval detector should operate on timing information alone. It was apparent from the experiments in that chapter that the squared-envelope detector performs better under some conditions (near centre detection), and the interval detector performs better under others (off-centre detection). This chapter developed a joint interval-peak detector which, when given both pieces of information together, outperforms, or at least matches the performence of, both the squared-envelope and interval detectors.

Finally, whereas the previous chapter had only considered detection using a single zero crossing interval received in one channel, this chapter tackled the problem of combining the information from many intervals. Attempts to extend the elementary interval detector to process multiple intervals led to the need for accurate and rapid evaluation of high-dimensional orthant probabilities; unfortunately, there is no technique available to do this at the present time. Instead, a (single-channel) multiple interval detector was developed using conventional post-integration methods: independent zero crossing intervals were sampled at regular intervals and averaged to form a Gaussian statistic.

Chapter

6

Tracking and Grouping Tonals in the ZCPA

The purpose of this chapter is somewhat broader than those of earlier chapters, as it moves beyond timing-based *detection* (e.g., deciding whether a signal is present using zero crossings) to consider *estimation*, i.e, the problem of determining an unknown signal quantity; *tracking*, which is the extension of estimation over time; and *grouping*, i.e., deciding whether two or more tracks indicate a common source. There is also an opportunity to re-examine the ZCPA presented in Chapter 3 in light of the theoretical work on zero crossing intervals immediately above. The chapter draws to a close many aspects of the preceding work, as well as opening up some new directions to explore, which inevitably results in a change of emphasis.

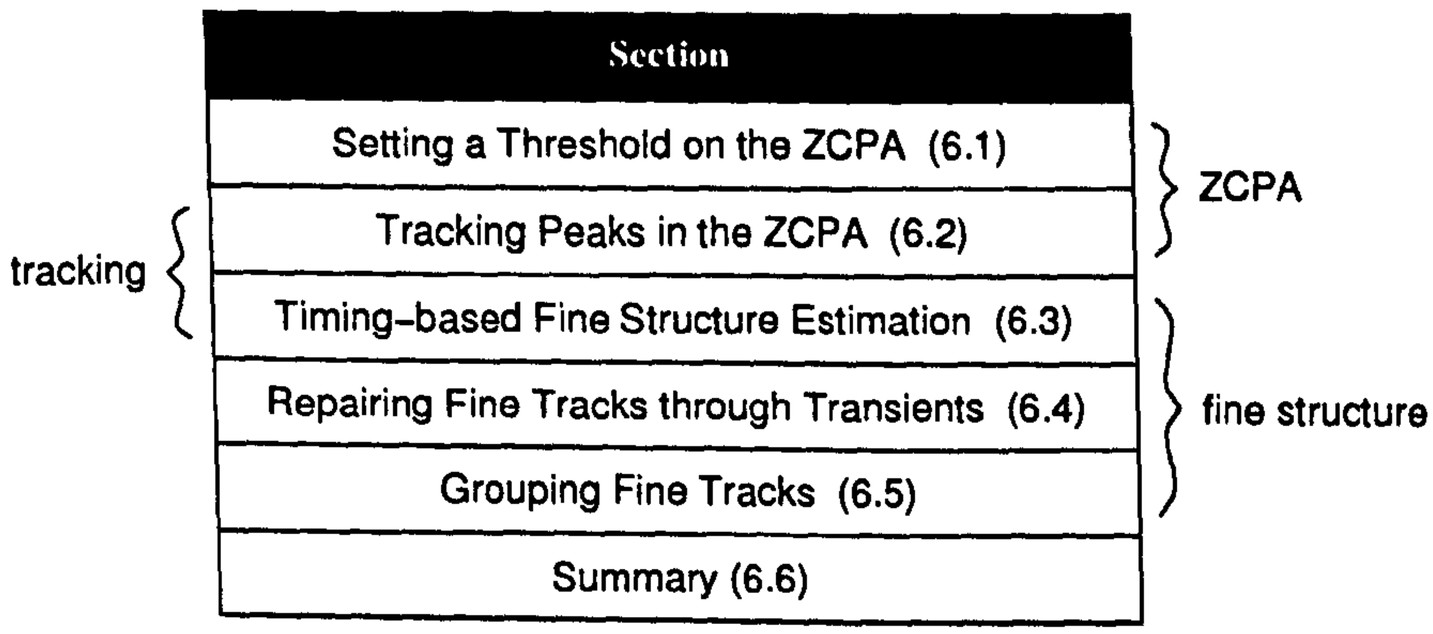
First, there is a focus on *reconciliation*: an attempt to relate the rather abstract discussion of interval distributions, and the highly-controlled evaluation of model detectors using synthetic signals in Chapters 4 and 5, to the practical problem of deciding how to calibrate the ZCPA to best represent real-world sonar signals, over which we have little or no control. Several experiments have been conducted to determine the performance of optimal detectors in model scenarios, but a real sonar recording will contain numerous departures from ideal conditions. In many cases, even basic quantities such as overall signal gain, which heretofore have been taken for granted, will be unknown.

Second, there is a steady progression towards *object-orientation*. Auditory scene analysis is an attempt to describe how a listener, upon hearing a mixture of sounds, perceives whole "auditory objects" and groups them into streams (Bregman, 1990); but it is predicated upon, and constrained by, the findings of psychological hearing studies. One analogous goal of the following work is, by building on lower representational levels, to uncover simple elements in the raw sonar signal waveform, such as peaks, tracks and transients, and, if possible, to group them together on the basis of common features.

Third, the studies pay closer attention to *application*. Although using an elementary interval detector to detect a tonal in noise could be considered an application, in this chapter we will often speak in terms of the kind of activity a human sonar operator might perform: for example, "setting" a threshold for signal detection, "clicking" on a tonal track, "zooming in" on its fine structure, or "comparing" one track with another on a display.

Having overviewed the purpose, we turn now to the material of the chapter, which consists of five germinal studies, spanning three interlocking themes: the ZCPA, track formation and fine structure processing. The first study addresses the problem of placing a threshold on a ZCPA bin to decide whether or not a signal is present (§6.1). The second study examines whether, by applying this threshold on a frame-by-frame basis, it is possible to track signals that persist over time (§6.2). Having obtained a track, the third study investigates how one might estimate its time-varying frequency using timing-based methods (§6.3). The fourth study proposes a rudimentary transient detector and modifies the frequency estimation routine to dismiss unreliable measurements during transient events, such as knocks (§6.4). The fifth and final study describes a grouping algorithm for fusing frequency tracks that show a common pattern of modulation (§6.5).

Chapter 6 Outline



6.1 Setting a Threshold on the ZCPA

The DFT and ZCPA spectrograms are similar in that both consist of a two-dimensional array of cells spaced uniformly in time and frequency, and a higher value in a cell always makes the presence of a signal at that time and frequency more likely, albeit via different mechanisms. There are well-established distributions governing the DFT magnitude for simple classes of signal and noise, such as Gaussian processes and sinusoids. Ultimately, these are derived by considering how individual time-domain samples with simple statistical behaviour are combined in the DFT processor. We must now ask, "Can a similar philosophy be used to derive the distribution of a ZCPA bin?"

The value stored in a ZCPA bin is a random variable found by summing the contribution of intervals across a block of filters. Let us consider the ZCPA bin indexed k. The set $\mathcal{S}_k$ contains only the indices of the filters that are capable of contributing to bin k. The set $\mathcal{I}$ consists of the indices of the most recent interval-peak pairs used to form the histogram. As the ZCPA bin value is derived from many, *individual* intervals, we shall attempt to draw on the earlier work regarding the distribution of a single interval (and its peak value) as a route to finding the mean, variance, and possibly higher moments, of the ZCPA bin.

Mean

Let $C_k(\iota, s)$ denote a *contribution function*, that is, the amount by which the ι-th interval of the s-th channel increases the k-th histogram bin. (The specific interpretation of this function will become apparent shortly.) In this case, the mean value of ZCPA bin k is given by

$$E\{\text{ZCPA}[k]\} = E\left\{\sum_{s\in\mathcal{S}_k}\sum_{\iota\in\mathcal{I}} C_k(\iota, s)\right\} \tag{6.1}$$

$$= \sum_{s\in\mathcal{S}_k}\sum_{\iota\in\mathcal{I}} E\left\{C_k(\iota, s)\right\} \tag{6.2}$$

$$= |\mathcal{I}| \sum_{s\in\mathcal{S}_k} E\left\{C_k(s)\right\}. \tag{6.3}$$

Having assumed that the input signal is stationary, the expected value of an interval is independent of its index, so step (6.3) makes use of $E\{C_k(\iota, s)\} \equiv E\{C_k(s)\}$. (We write $|\mathcal{I}|$ for the cardinality of the set $\mathcal{I}$.)

Variance

The derivation of the bin mean above essentially reduced the expectation of a sum to the sum of an expectation, and the same approach can be attempted to find the bin variance. The variance is obtained by subtracting the square of the mean from the

second raw moment. The second raw moment is written as follows.

$$E\{\text{ZCPA}^2[k]\} = E\left\{\left(\sum_{s\in\mathcal{S}_k}\sum_{\iota\in\mathcal{I}} C_k(\iota,s)\right)^2\right\} \tag{6.4}$$

$$= \sum_{s_1\in\mathcal{S}_k}\sum_{\iota_1\in\mathcal{I}}\sum_{s_2\in\mathcal{S}_k}\sum_{\iota_2\in\mathcal{I}} E\left\{C_k(\iota_1,s_1)C_k(\iota_2,s_2)\right\}. \tag{6.5}$$

The summand, $E\left\{C_k(\iota_1,s_1)C_k(\iota_2,s_2)\right\}$, expresses the raw covariance of interval contributions measured at different times and in different channels.

There appear to be severe difficulties facing the evaluation of these covariances, except in the trivial cases, in which either (i) the indices s_1, s_2, ι_1 and ι_2 conspire to make the intervals independent, in which case

$$E\left\{C_k(\iota_1,s_1)C_k(\iota_2,s_2)\right\} = E\left\{C_k(s_1)\right\}E\{C_k(s_2)\},$$

or (ii) the raw covariances are in fact *raw second moments*, i.e., $s_1 = s_2, \iota_1 = \iota_2$. An expression for the multiple interval probability distribution is required to evaluate these expectations explicitly, and as we observed in the previous chapter, a convenient, analytical form of this distribution is lacking at present. It is notable that even the specific problem of determining the joint distribution for *two successive* intervals of a Gaussian process—disregarding of the peak amplitude—has received attention as a problem in its own right (Rychlik, 1987).

6.1.1 Mean Noise Profile of the Timing-only ZCPA

The *mean profile* of the ZCPA refers to the shape produced by averaging the ZCPA over a theoretically infinite number of rows; or, stated another way, it is a graph plotting the expected value of a bin against the bin frequency. The mean profile given noise-only conditions is a useful guide for placing an absolute threshold for signal detection, as the addition of a signal on the bin centre always raises its average value.

The preceding section introduced the contribution function, $C_k(\iota,s)$, to describe the increase in ZCPA bin k following the arrival of the interval ι in channel s. Let us consider the bin k, depicted in Figure 6.1, with lower and upper edges denoted f_{k0} and f_{k1}, respectively. A bin in the timing-only ZCPA is incremented by one, if and only if the time difference between two consecutive zero crossings in the same direction, i_p, falls into the bin's range, i.e.,

$$C_k(\iota,s) = \begin{cases} 1, & f_{k0} \le \dfrac{1}{i_p(\iota,s)} < f_{k1} \\ 0, & \text{otherwise.} \end{cases} \tag{6.6}$$

Exploiting the fact that intervals in a narrowband channel vary slowly, we can assume that $i_p \approx 2i_c$. The expected value of the contribution function can then be written in terms of the cumulative distribution function governing i_c, that is,

$$E\{C_k(s)\} = P\left(i_c \le \frac{1}{2f_{k0}} \text{ given } s\right) - P\left(i_c \le \frac{1}{2f_{k1}} \text{ given } s\right). \tag{6.7}$$

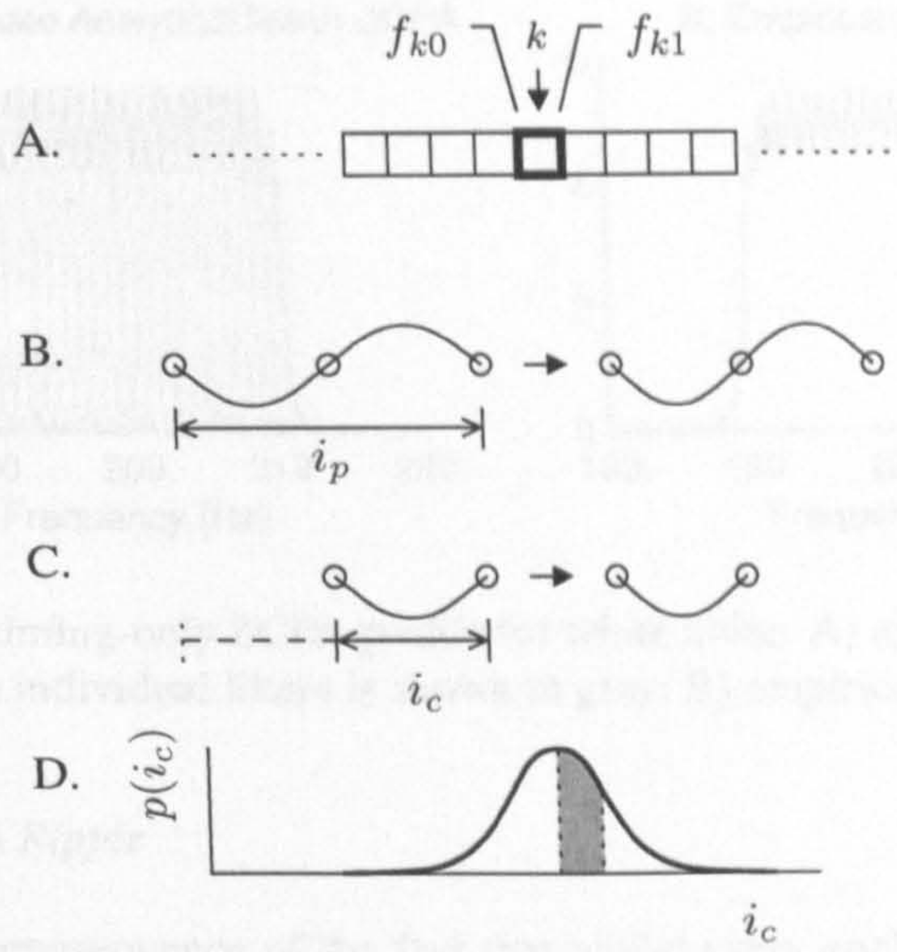


Figure 6.1: Deriving the contribution function for the timing-only ZCPA from the interval distribution. A) The expected contribution of an interval to bin k is the same as B) the probability that the interval between two upward zero crossings (i_p) falls into the bin range, which is approximately equal to C) the probability that twice the interval between two successive zero crossings (i_c) falls into the bin range, which can be found by D) differencing the cumulative interval distribution function.

A suitable approximation for the interval c.d.f. of a wide-sense stationary Gaussian process was derived in Chapter 4. Applying this in (6.7) and then (6.3), we can find the mean ZCPA spectrum for a white or coloured Gaussian noise signal.

Figure 6.2A shows the analytical mean profile of the timing-only ZCPA in response to white noise input. The analysis filters of the ZCPA are spaced 1 Hz apart, cover a range of 190 Hz–210 Hz, have tuning parameter $\alpha_a = 2.5$, and each contribute 20 intervals. The histogram bins are 0.25 Hz wide. An empirical mean profile generated from five thousand seconds of a white noise signal is provided in Figure 6.2B for comparison. The slight differences between the analytical and empirical profiles can be dismissed as the result of the accumulation of approximation errors[1]; the overall similarity between the two is sufficiently compelling at this stage.

More interestingly, the ZCPA profile itself (whether analytically or empirically derived) fluctuates around a steady state of 5 intervals. This is the desired response for white noise signal, as every filter contributes 20 intervals, and the filters and histogram bins are in a 4:1 ratio. However, the spiky fluctuation around the steady state, which we shall henceforth term *ZCPA ripple*, is a poor representation of a white noise spectrum, and we must devote a short section to discussing means to counteract its appearance.

[1] Such errors arise from the truncation of the impulse response, the interpolation of zero crossings, the (false) assumption of perfect interval conditioning when computing the interval p.d.f., and so on.

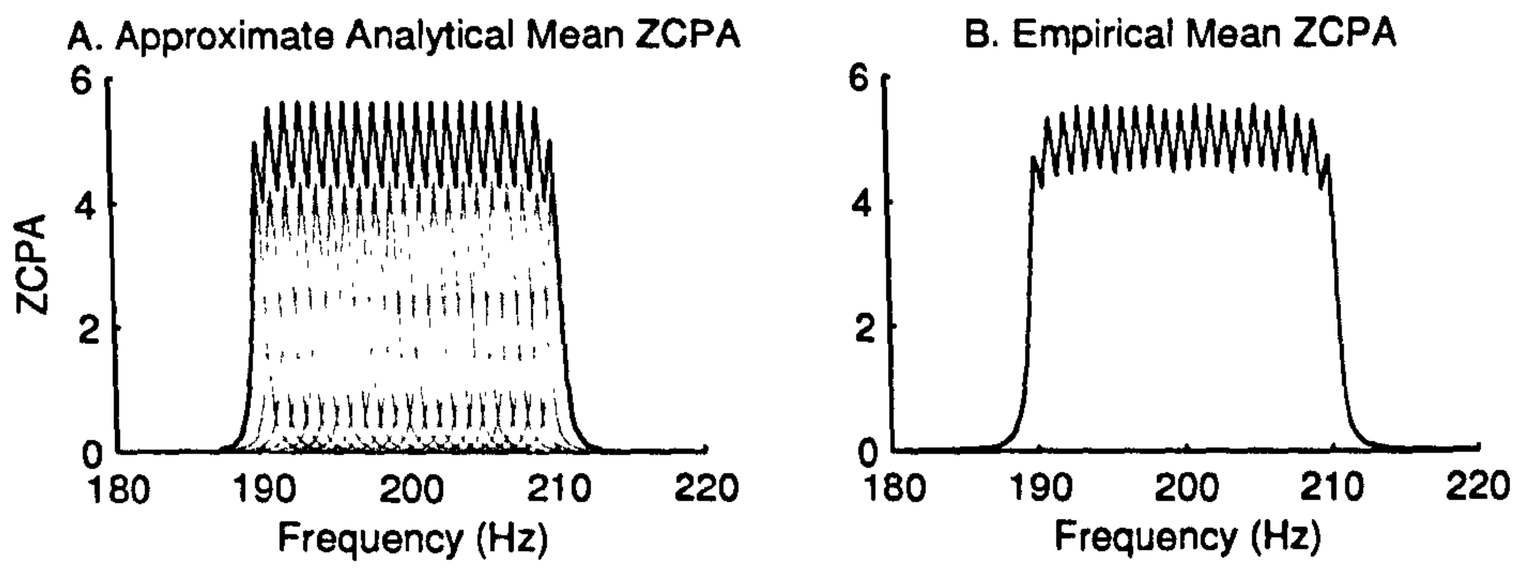


Figure 6.2: Mean timing-only ZCPA profile for white noise. A) analytical profile, with contributions from individual filters is shown in grey; B) empirical profile.

Timing-only ZCPA Ripple

ZCPA ripple is a consequence of the fact that whilst every analysis filter produces a range of zero crossing intervals roughly commensurate with its magnitude response, each has an increased tendency to output intervals nearer to its centre frequency. The mean contributions of individual analysis filters are plotted in Figure 6.2A as a series of grey 'spikes'; the final mean profile is found by adding these together. Evidently the (reciprocal) interval distribution of every analysis filter is decidedly sharper than its squared-magnitude response, which is Gaussian-shaped, and the ZCPA bins which coincide with filter centres draw more intervals than those those which fall inbetween.

The most natural solution to the ripple problem is to increase the bandwidth of the analysis filters so that their respective interval distributions grow wider and sum to form a flatter profile. We may recall from earlier chapters that the filter bandwidth is controlled by two parameters: the impulse response duration, T_a, and the tuning parameter, α_a. When modelling a theoretical Gaussian window of infinite duration, the two parameters combine to form a single parameter, $C = \alpha_a / T_a$. In the DFT, however, T_a is fixed according to the DFT length, and α_a controls the sharpness of the window function. In general, larger values of α_a correspond to wider bandwidths.

Having traced the cause of the ripple and noted the relevant parameters, we shall select a few values for α_a and monitor their effect upon the ZCPA profile. Figure 6.2 above was produced by setting $\alpha_a = 2.5$; Figures 6.3A–C show the profiles for $\alpha_a = 2.0$ (narrower bandwidth), and $\alpha_a = 4.0$ and $\alpha_a = 6.0$ (wider bandwidths), respectively. The suppression of ZCPA ripple for larger values of α_a is apparent from these figures and supports our earlier reasoning. It should be remembered that, because α_a controls analysis bandwidth, it has an impact on the signal-to-noise ratio and resolution of the ZCPA, in addition to ripple in the mean profile. Although one could, in principle, choose a very large value of α_a to suppress ripple almost entirely, in practice, it would also lower the post-analysis SNR and lead to interference between closely-spaced components.

In order to establish a suitable trade-off, we shall quantify ripple as a function of α_a and then attempt to find the smallest value of α_a for which the level of ripple is tolerable.

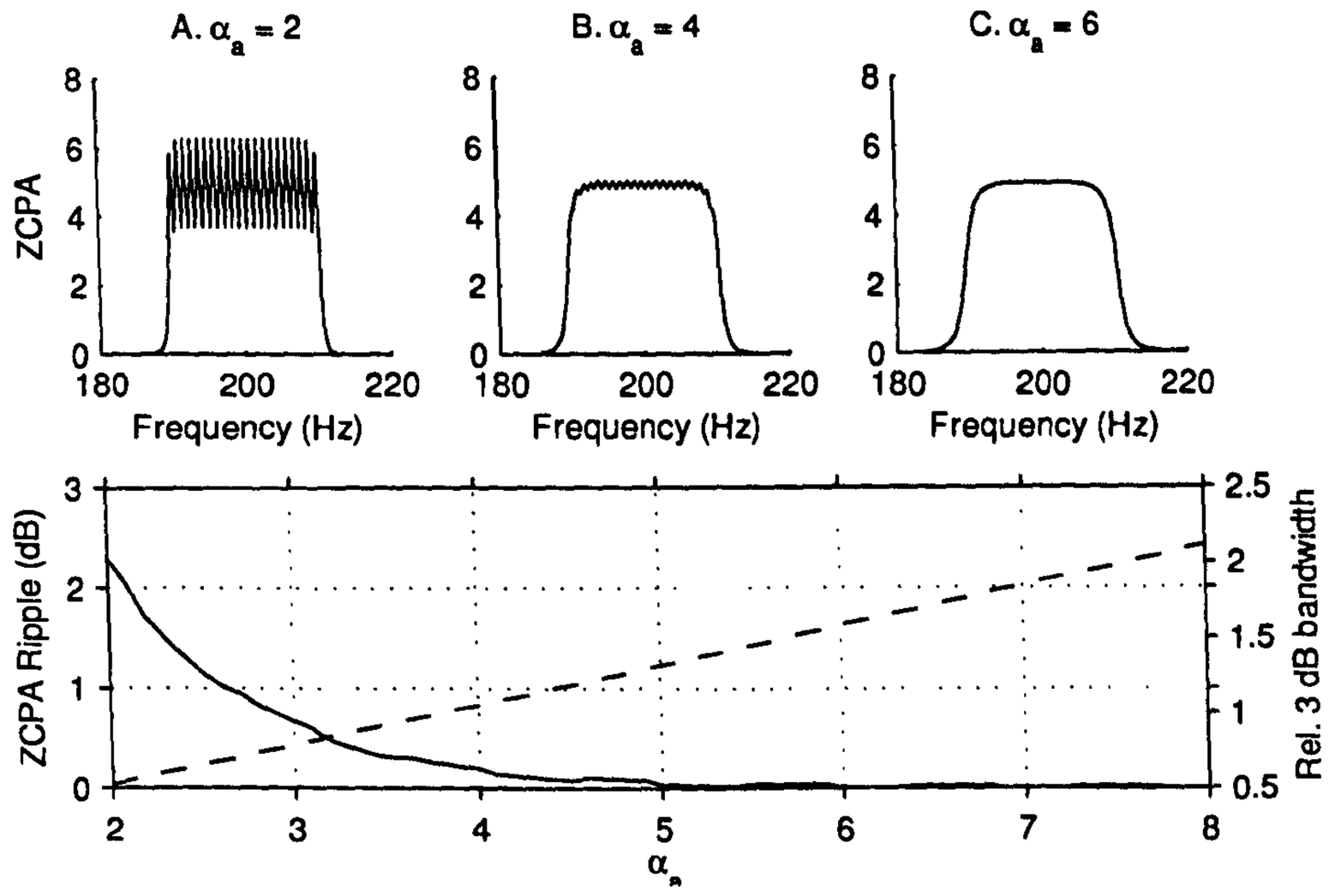


Figure 6.3: The upper panels plots the mean timing-only ZCPA profile for various α_a. The lower panel plots ripple factor (dB, solid line) and relative bandwidth (ratio of 3 dB bandwidth to binwidth, dashed line) as a function of α_a.

Ripple in the band-pass portion of a filter magnitude response is usually measured using the decibel ratio of peak to trough levels (Oppenheim and Schafer, 1989). Adopting a similar approach here, the lower panel of Figure 6.3 writes the ripple factor (in dB) on the left-hand ordinate, the scaled filter bandwidth on the right-hand ordinate, and α_a on the abscissa. From these curves, it appears that $\alpha_a = 4.0$ is a good choice, as it roughly matches the 3 dB bandwidth with the DFT bandwidth and leaves only a very small amount of ripple (see also Figure 6.3B); at the expense of a small increase in bandwidth, $\alpha_a = 5.0$ suppresses ripple almost entirely; there is negligible ripple for all $\alpha_a > 5.0$.

Figure 6.4 shows the timing-only ZCPA for one minute of sonar recording, computed with $\alpha_a = 4.0$, and plots the mean profile underneath. A flat profile of five intervals is apparent in regions of the noise, even though the energetic noise floor is certainly not flat (*cf.* Fig. 3.4); dominant components appear as spikes of varying height, flanked by small troughs. (The significance of this shape is discussed in Section 6.1.3.) The appearance of the profile suggests placing a detection threshold at ten intervals; this is marked on the figure as a dotted line. The key result here is that, in the case of the timing-only ZCPA, the expected height of the noise floor is known in advance, being determined from the ZCPA configuration rather than the data.

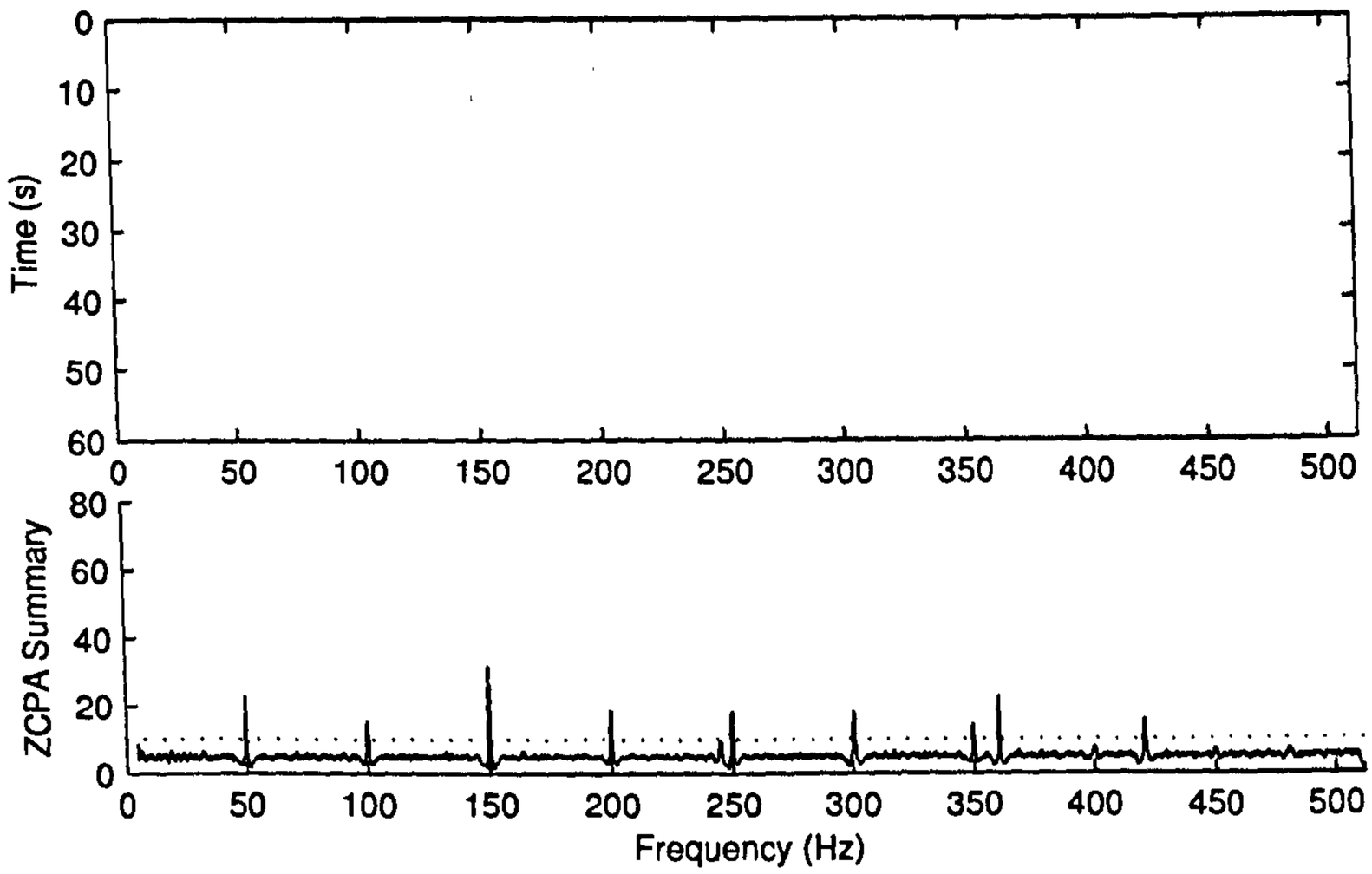


Figure 6.4: ZCPA and mean ZCPA profile for a minute-long recording of an oil tanker. (See Figure 3.19 and caption.)

6.1.2 Mean Noise Profile of the Peak Squared Amplitude ZCPA

The standard ZCPA weights the contribution of an interval to the histogram according to some nonlinear compression of its peak amplitude. The contribution function of the timing-only ZCPA only accounted for the probability that an interval is assigned to a particular bin. For the peak squared amplitude ZCPA, we must consider not only the probability that (i) an interval is assigned to given bin, but also (ii) the expected value of its peak squared amplitude, *given* the bin to which it is assigned. Using $E(\iota, s)$ to denote the peak squared amplitude for the interval indexed ι in channel s, the contribution function is modified accordingly.

$$C_k(\iota, s) = \begin{cases} E(\iota, s), & f_{k0} \leq \dfrac{1}{i_p(\iota, s)} < f_{k1} \\ 0, & \text{otherwise.} \end{cases} \tag{6.8}$$

The expected value of this function can be found by integration the joint p.d.f., i.e.,

$$E\{C_k(\iota, s)\} = \int_{1/(2f_{k1})}^{1/(2f_{k0})} \int_0^\infty p_{I_c E}(i_c, e)\, de\, di_c \tag{6.9}$$

$$= \int_{1/(2f_{k1})}^{1/(2f_{k0})} \left[\int_0^\infty p_{E|I_c}(e \mid i_c)\, de \right] p_{I_c}(i_c) di_c. \tag{6.10}$$

The expression in squared brackets is the expected value of the squared amplitude, E, given that the interval over which it occurs has duration i_c. We can evaluate the inner

integral using the approximation for $p_{E|I_c}(\cdot)$ given in (5.77), i.e.,

$$\int_0^\infty \tilde{p}_{E|I_c}(e \mid i_c)de = \int_0^\infty \sqrt{\frac{e}{2\pi|\Sigma|^3}} \exp \frac{e}{-2\Sigma}\, de = 3\Sigma, \tag{6.11}$$

where

$$\Sigma = \frac{\gamma_X^2(0) + \gamma_X(0)\gamma_X(i_c) - 2\gamma_X^2(\frac{1}{2}i_c)}{\gamma_X(0) + \gamma_X(i_c)}. \tag{6.12}$$

Then, placing (6.11) and (6.12) into (6.10), we get

$$E\{C_k(\iota, s)\} = 3 \int_{1/(2f_{k1})}^{1/(2f_{k0})} \left(\frac{\gamma_X^2(0) + \gamma_X(0)\gamma_X(i_c) - 2\gamma_X^2(\frac{1}{2}i_c)}{\gamma_X(0) + \gamma_X(i_c)} \right) p_{I_c}(i_c)\, di_c. \tag{6.13}$$

It will be difficult obtain a closed-form solution to this integral. However, because the bins are relatively narrow, we can assume that both i_c and $E\{E \mid i_c\}$ change very gradually over the region of integration. One possible way forward involves expanding the integrand using a first- or second-order Taylor series; another is to assume that the expected value of the envelope varies so slowly with i_c that it can be replaced with a constant equal to its value at the mid-point of the integral[1]. The latter approach produces the rather ungainly but conceptually simple expression,

$$\begin{aligned} E\{C_k(\iota, s)\} \quad = \quad & 3 \left[\frac{\gamma_X^2(0) + \gamma_X(0)\gamma_X\left(\frac{f_{k0}+f_{k1}}{4f_{k0}f_{k1}}\right) - 2\gamma_X^2\left(\frac{f_{k0}+f_{k1}}{8f_{k0}f_{k1}}\right)}{\gamma_X(0) + \gamma_X\left(\frac{f_{k0}+f_{k1}}{4f_{k0}f_{k1}}\right)} \right] \\ & \times \left[P\left(i_c \le \frac{1}{2f_{k0}} \text{ given } s\right) - P\left(i_c \le \frac{1}{2f_{k1}} \text{ given } s\right) \right]. \end{aligned} \tag{6.14}$$

All the quantities in (6.14) are known, and the mean ZCPA profile can now be found.

Figures 6.5A and 6.5B respectively plot the analyical and empirical mean ZCPA profile for a white Gaussian noise signal with unit power spectral density. The analysis filters and histogram bins are the identical to those used for the timing-only ZCPA study above, and $\alpha_a = 2.5$. As before, the dissimilarity between the two profiles can be attributed to cumulative approximation errors (in particular, it seems, the truncation of the impulse response for small α_a).

Peak Squared Amplitude ZCPA Ripple

The lower panel of Figure 6.6 plots the ripple factor in the peak squared amplitude ZCPA profile (that is, the decibel peak-to-trough ratio) as a function of α_a. As with the timing-only ZCPA, the ripple in the mean peak squared amplitude ZCPA profile

[1] This is tantamount to replacing the first part of the integral with a zeroth-order Taylor series expansion around the centre interval,

$$\frac{1}{2}\left(\frac{1}{2f_{k0}} + \frac{1}{2f_{k1}}\right) = \frac{f_{k0} + f_{k1}}{4f_{k0}f_{k1}}.$$

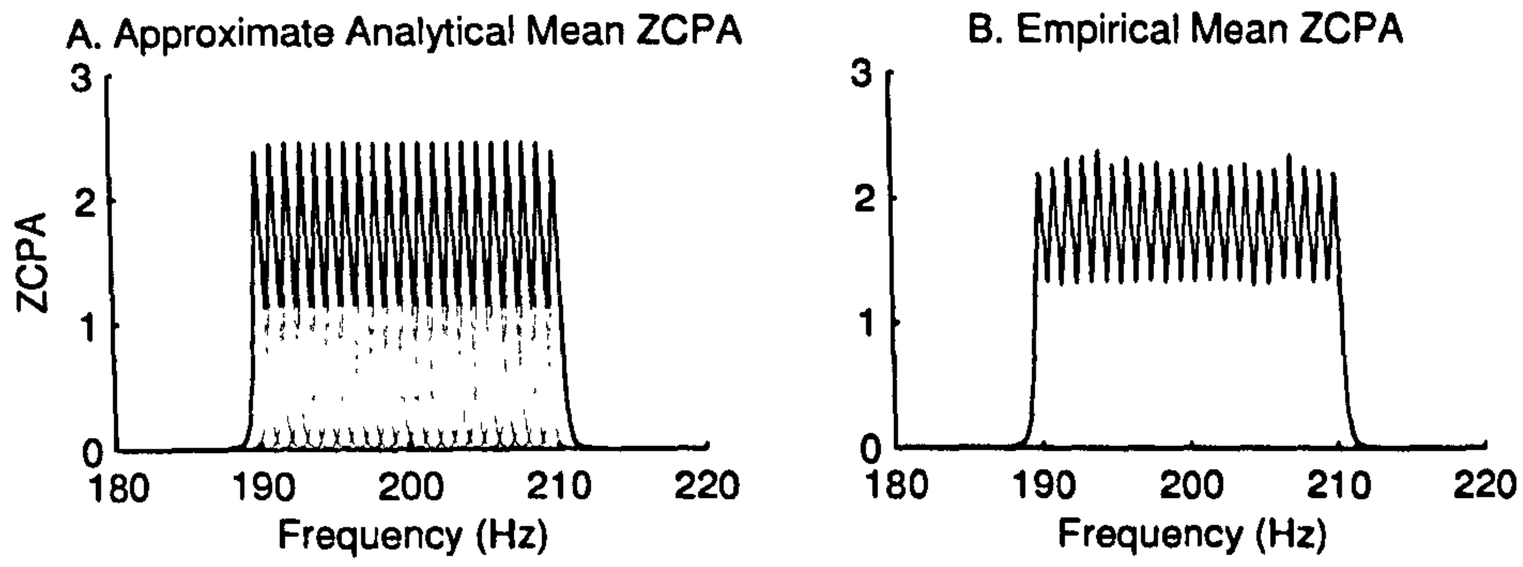


Figure 6.5: Mean peak squared amplitude ZCPA profile for white noise. A) analytical profile, with contributions from individual filters is shown in grey; B) empirical profile.

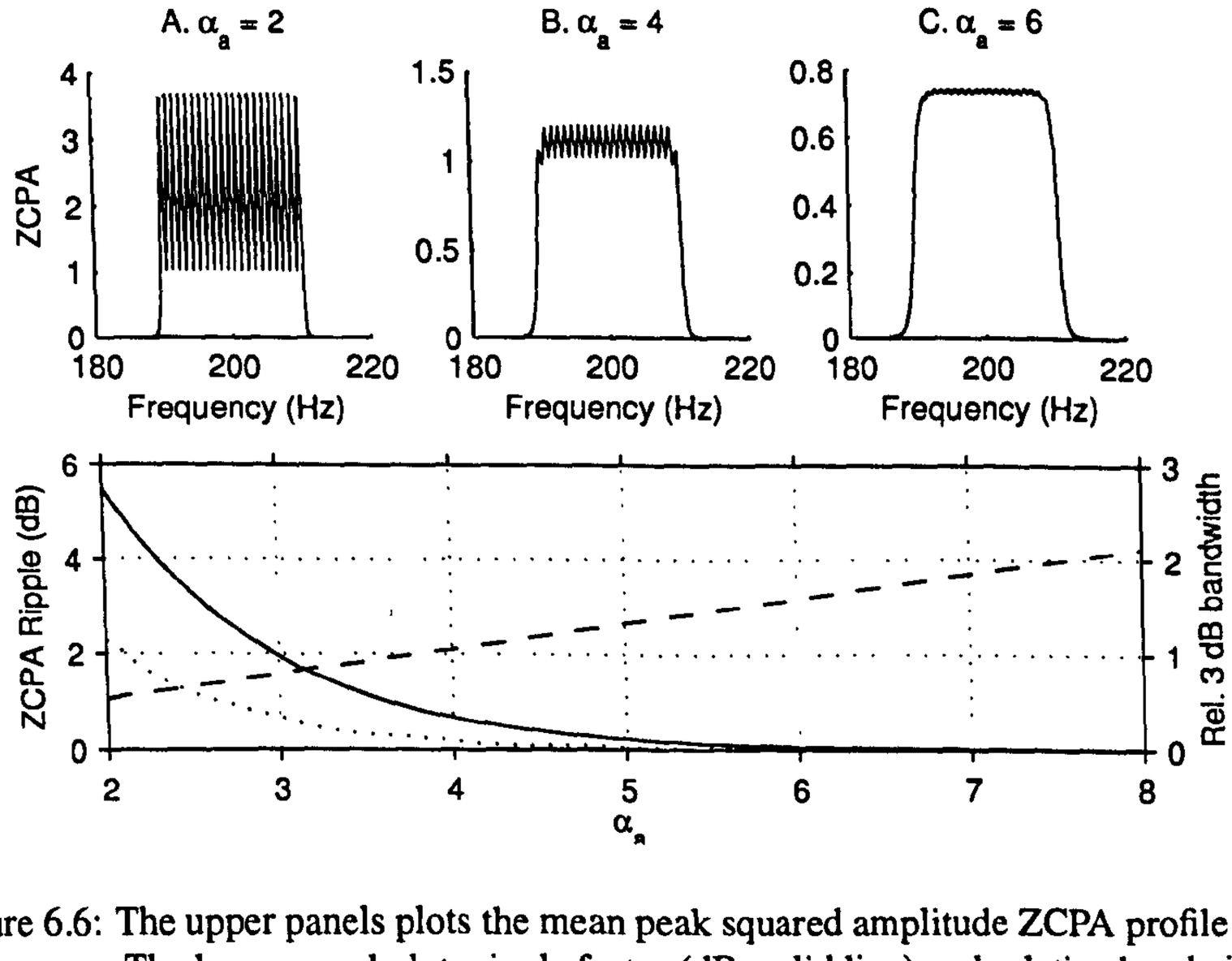


Figure 6.6: The upper panels plots the mean peak squared amplitude ZCPA profile for various α_a. The lower panel plots ripple factor (dB, solid line) and relative bandwidth (ratio of 3 dB bandwidth to binwidth, dashed line) as a function of α_a. The ripple factor of the timing-only ZCPA is included for comparison (dotted line).

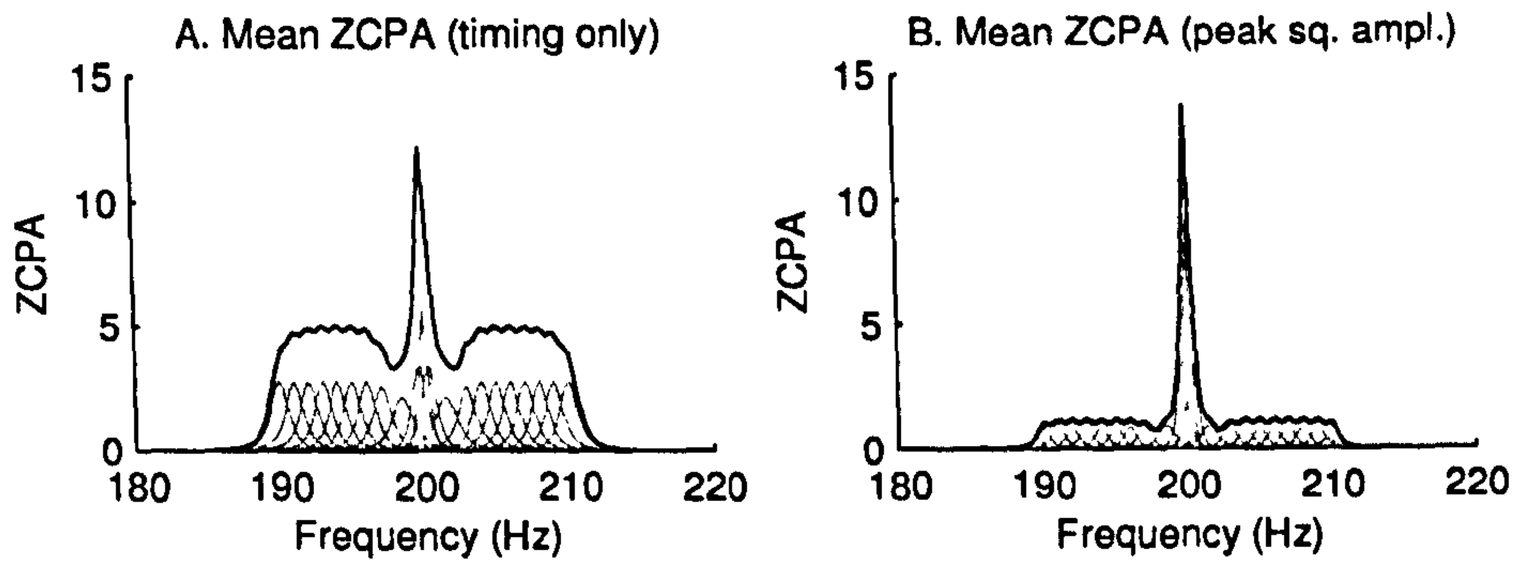


Figure 6.7: A) mean timing-only ZCPA profile and B) mean peak squared amplitude ZCPA profile for a narrowband signal in noise, with components plotted in light grey.

is due to the tendency of the analysis filters to output intervals close to their centre frequencies when driven by white noise. Increasing α_a broadens the response of the analysis filters and the associated interval probability density functions, thus flattening the average ZCPA response. It should also be noted that, unlike the timing-only ZCPA, the increase in attenuation due to larger α_a rescales the mean profile.

Figure 6.6 plots the ripple factor of the peak squared amplitude profile against that of the timing-only profile with a solid and dotted line, respectively. A comparison of the curves reveals that the ripple factor is worse in the former than in the latter. As we have noted, an analysis filter contributes intervals near its centre frequency more often (due to the dominant frequency principle); however, the intervals near the centre are also weighted by larger peak amplitudes, as there is less attenuation in the band near its centre. This combination of effects means that larger value of α_a is required to match the ripple factor of the peak squared amplitude ZCPA against that of the timing-only ZCPA. A value of $\alpha_a = 5.0$ appears to be a sensible compromise.

6.1.3 Mean Signal-and-Noise Profile of the ZCPA

Although only white noise processes have been considered so far in this section, we are able, in principle, to calculate the mean ZCPA profile for a number of signal and noise configurations, including additive mixtures of sinusoidal, notched noise and coloured noise processes. In drawing this topic to a close, we shall examine the mean ZCPA profile of just one type of mixture: the narrowband signal in noise configuration which featured in Chapter 4.

The signal process consists of white noise with unit power spectral density convolved with an impulse response whose MGMM description is

$$\begin{aligned}\Lambda_{h_s} &= \langle A = 2.5, C = 0.625, \mu = 0, \bar{\omega} = 2\pi \cdot 200, \phi = 0\rangle \\ &\quad + \langle A = 2.5, C = 0.625, \mu = 0, \bar{\omega} = -2\pi \cdot 200, \phi = 0\rangle ,\end{aligned}$$

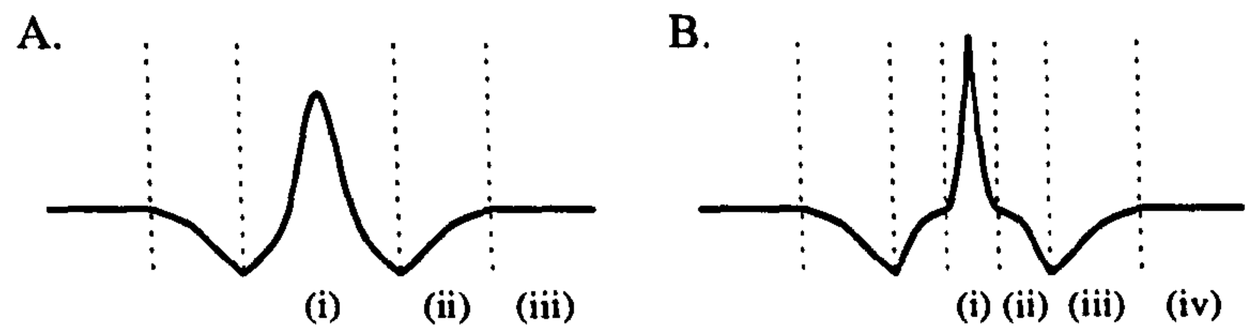


Figure 6.8: A) synchronised regions in the timing-only ZCPA: (i) synchrony excess; (ii) synchrony deficit; (iii) no synchrony (noise floor); B) energised and synchronised regions in the peak squared amplitude ZCPA: (i) synchrony excess and energy excess; (ii) energy excess and synchrony deficit; (iii) synchrony deficit and no energy excess; (iv) no synchrony and no energy excess (noise floor).

This is added to a white noise background with unit power spectral density, giving a narrowband signal-to-noise ratio[1] of approximately 9.5 dB.

The mean timing-only ZCPA profile ($\alpha_a = 4.0$) is shown in Figure 6.7A. It is clear from the mean profile components, shown in light grey, that the spectral dominance at 200 Hz has drawn intervals from the surrounding channels. The intervals of the (200 ± 1) Hz channels gravitate signficantly away from their own centres, towards the signal frequency; those of the (200 ± 2) Hz channels are moderately affected; those of the (200 ± 3) Hz channels are negligibly affected; and beyond this, the signal exerts no influence. The intervals of the adjacent filters that are captured by the signal do not contribute to their "local" bins, and, consequently, two grooves are etched in the mean profile either side of the peak.

The mean peak squared amplitude ZCPA profile ($\alpha_a = 4.0$), shown in Figure 6.7B, exhibits a small amount of synchrony capture. In the locations where the grooves would ordinarily appear in the timing-only ZCPA, there is a small bump, due to the fact that although the neighbouring filters do not capture as many intervals, the intervals they do capture also receive some energy from the component. The unusual shape of the peak squared amplitude ZCPA profile is caused by an *interaction* between energetic and synchronised regions—an interaction which is not present in the timing-only ZCPA, or indeed the "energy-only" DFT. A speculative attempt to label the characteristic features of the ZCPA profiles is related in Figure 6.8.

The mean ZCPA noise profile and the shape of the synchrony profile—both of which we must recall are derived from the interval (and joint interval-peak) distribution in earlier chapters—will to some extent inform the design of the peak tracking routines that follow.

[1] The narrowband SNR was defined as the dB ratio of total signal power to noise power in a 1 Hz band.

6.2 Tracking Peaks in the ZCPA

The automated detection of narrowband signals in the ZCPA can be accomplished on a frame-by-frame basis using a suitably-chosen threshold, perhaps one placed at some multiple of the mean ZCPA noise profile. Figure 6.9A shows the result of applying a threshold to the timing-only ZCPA displayed in Figure 5.4: values exceeding fifteen are mapped to black pixels; the rest are mapped to white. (The mean noise profile is five.) The threshold highlights the tonal peaks at the cost of introducing many thousands of spurious peaks. One principle for discerning whether a peak is genuine—and also one of the Gestalt principles mentioned in Chapter 2—is *continuity*, that is, the persistence (or transience) of a peak over time.

The final problem to be tackled here is that of distilling from the two-dimensional grid of "peak candidates" a set of continuous tonal tracks. This constitutes the first attempt in this work to move from a data-oriented representation to an object-oriented one, and it requires some care. Although the human eye is well-adapted to spot even the vaguest of structured features in an image (e.g., a line at 450 Hz in Fig. 6.9A), an automated peak tracker is susceptible to a number of faults. First, peaks frequently *disappear and reappear* as result of fluctuations in the noise floor or signal level. Similarly, peaks *vary in frequency* due to changes in the source, the influence of the channel, or simply as the result of noise. In addition, there may be *competing interpretations* of how to extend a track, e.g., should the algorithm opt for the nearest peak or the tallest peak? What happens if two tracks collide or cross?

Tracking algorithms of varying degrees of sophistication and auditory inspiration have been proposed in recent years and are often presented as one component in a larger system. Amongst those which track sinusoids (also called partials or harmonics) we include the CASA models of Cooke (1991/1993), Mellinger (1991) and Nakatani (2002), the spectral peak tracker of McAulay and Quatieri (1986), the HMM-based partial tracker of Depalle et al. (1993), the Kalman filter-based CASA system of Unoki and Akagi (1999) and the linear prediction-based model of Lagrange et al. (2004). Any of these solutions could, in principle, be applied to the problem of tracking tonal peaks in the ZCPA. In this work, we shall consider just one: the birth-death peak tracker of McAulay and Quatieri (1986).

6.2.1 Birth-Death Peak Tracking (McAulay and Quatieri)

The birth-death peak tracking scheme of McAulay and Quatieri (1986) can be described in high-level terms as follows. The routine maintains a set of time-frequency tracks, which are created, extended and terminated in an online fashion. A set of peaks is determined for each spectral frame on its arrival and then processed. If a peak is sufficiently close to the end-point of a "live" track, the track is extended accordingly and the peak is removed from the set. The remaining peaks, which cannot be joined to a track, each give "birth" to a new, live track. Conversely, the inability of any live track to perpetuate itself by connecting to an arriving peak causes its "death".

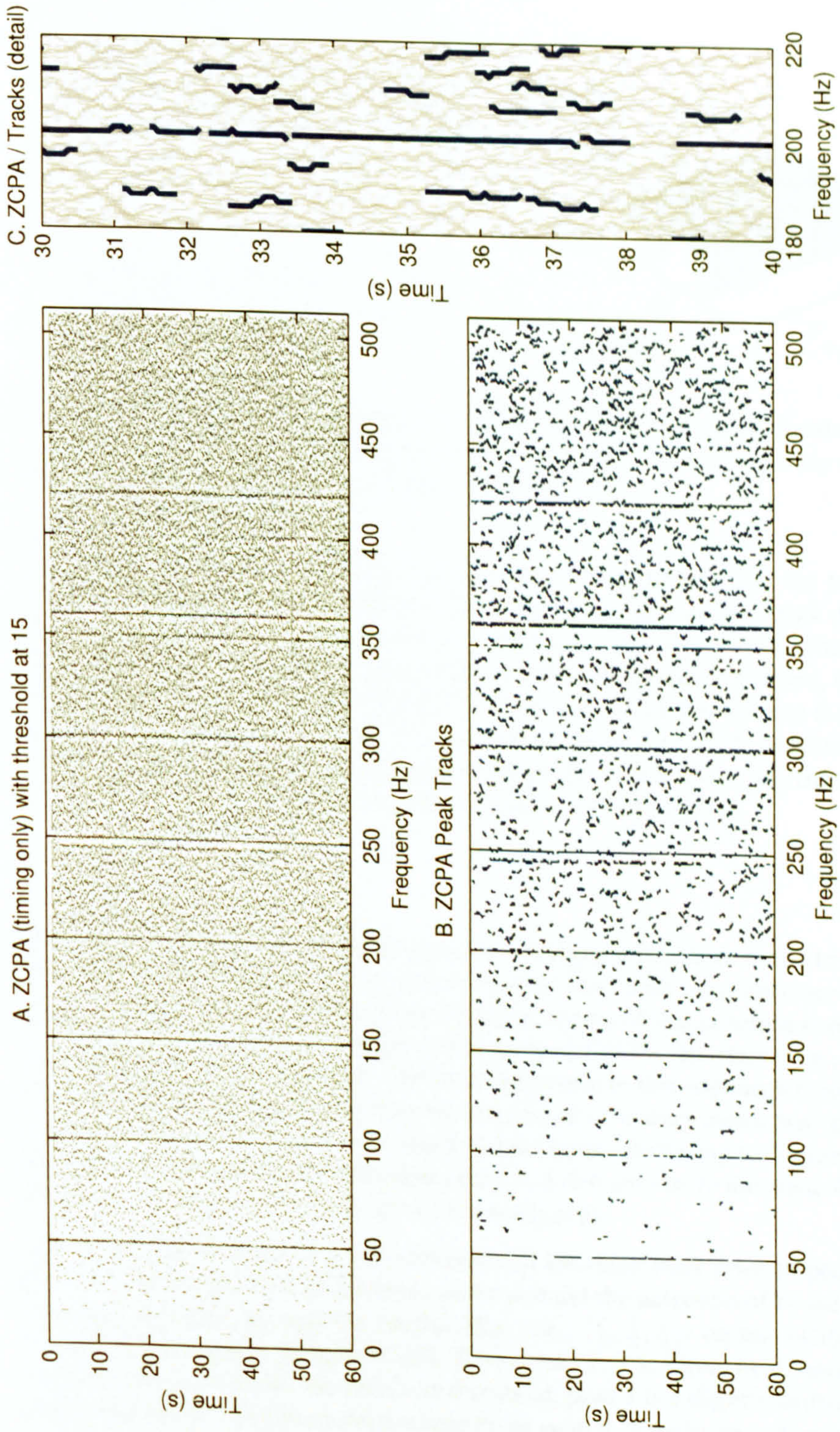


Figure 6.9: A) a constant threshold applied to the ZCPA time-frequency plane; B) the birth-death peak tracking scheme of McAulay and Quatieri (1986) applied to the ZCPA output; C) a magnified section of (B) atop (A). All plots are based on the same merchant vessel recording.

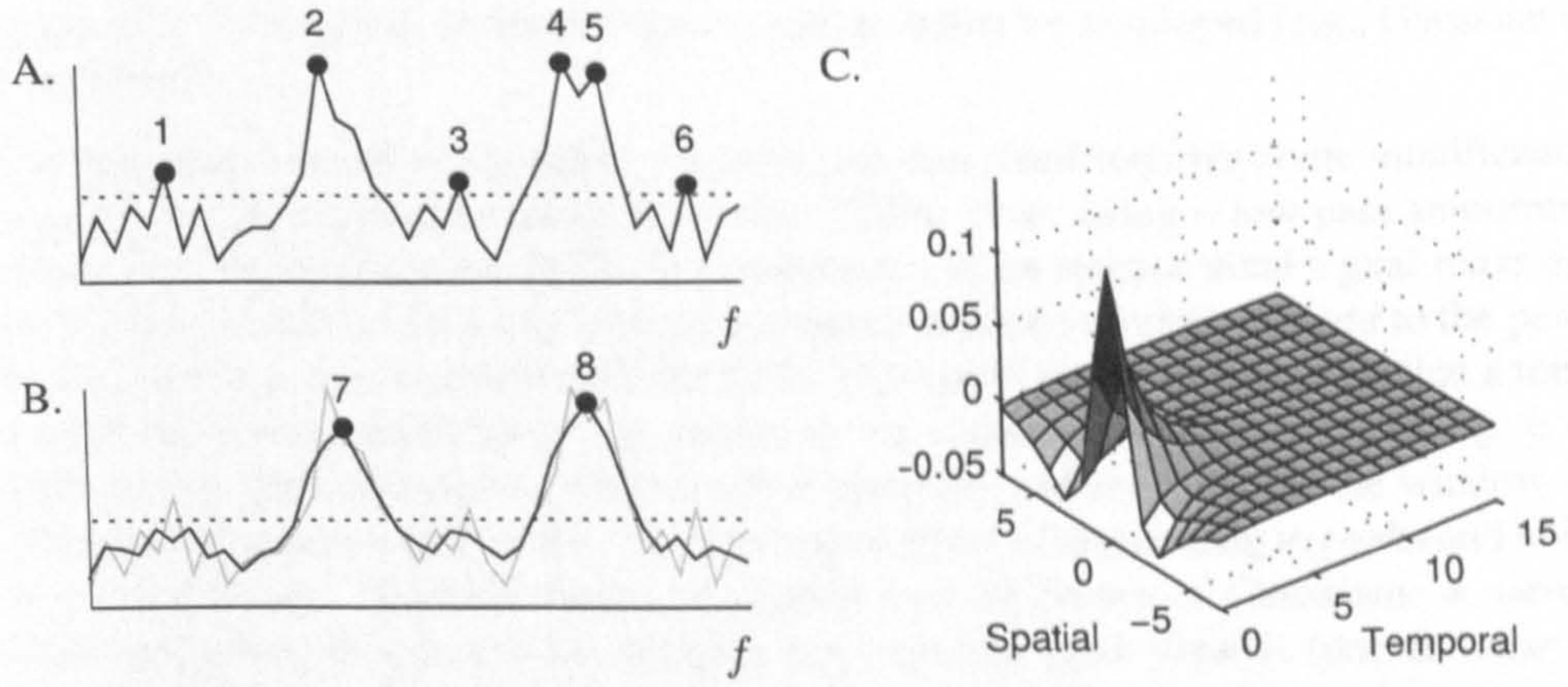


Figure 6.10: Simple peak detection. A) peaks in the ZCPA above threshold (dashed line); B) peaks in the smoothed ZCPA above threshold. C) time-frequency convolution kernel used in the production of Figures 6.9B and 6.9C.

The McAulay-Quatieri algorithm is well-suited to our present task for two reasons. First, each module of the algorithm—that is, the spectral analyser, peak detector and peak tracker—is *interchangeable*, in the sense that it can be replaced with a different implementation without requiring changes in the other modules. Second, the modules function in a strictly *feed-forward* and *online* fashion, which leaves open the possibility of tracking peaks in the ZCPA as it is processed. Here, the ZCPA constitutes the first module of the three-module architecture (rather than the Fourier spectrum); we shall now describe the implementation of the second and third modules.

Peak Detection

The most direct method for detecting peaks is to apply a simple criterion to each ZCPA bin, labelling it as a peak if it is strictly greater than both its immediate neighbours and the detection threshold. The peak detection aspect of this criterion is equivalent to finding negative-going sign changes in the spectral / ZCPA derivative (i.e., a transition from positive to negative slope). The main objection to this approach is its sensitivity to spurious peaks above the detection threshold and secondary peaks that occur in what a human viewer would label as a single, broad peak. Referring to Figure 6.10A, for example, an observer might judge that peaks 1, 3 & 6 are due to the noise floor, peak 2 is genuine, and peaks 4 & 5 are part of the same peak.

Convolving the ZCPA with a two-dimensional low-pass filter prior to peak detection can assist the suppression of spurious peaks and aid the detection of broad peaks. The effect of a three-point, uniform spatial filter (i.e., $\left[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right]$) on the ZCPA plotted in Figure 6.10A is shown in Figure 6.10B. Here, peaks 1, 3 & 6 have been smoothed to the extent that they fall under the detection threshold, peak 7 is a slightly shifted version of peak 2, and peaks 4 & 5 have been merged into peak 8. Simply smoothing the ZCPA is enough to address both problems cited above—spurious and divided peaks—although

in practice, a smoother, wider window would probably be employed (e.g., Gaussian or Blackman).

The procedure for detecting peaks we have just described requires some modification before it can be applied successfully to the ZCPA. First, using a low-pass smoothing window on the timing-only ZCPA is problematic, as an intense tonal signal *reassigns* intervals from the surrounding bins to its centre frequency; it does not *add* to the peak, as it would to a peak in the magnitude DFT. This could mean, for instance, that a tonal contributes twenty intervals to the centre of the Gaussian window but nothing to its sides, whilst the noise floor contributes five intervals uniformly across the window. In other words, such a window has the undesirable effect of responding to peaks and noise in equal measure. A better choice of window is a difference of Gaussians: a narrow Gaussian pulse, whose breadth matches the expected peak breadth (due to noise or signal bandwidth), from which is subtracted a wider Gaussian pulse, whose breadth matches the expected trough breadth (due to the capture of intervals). The shape of this spatial filter recalls that of the mean ZCPA profile shown in Figure 6.8A.

A second, lesser problem with spatial averaging is its capacity to move peaks. This is illustrated incidentally in Figure 6.10, where the unsmoothed peaks 2 and 4 correspond to the smoothed peaks 7 and 8—the latter shifted upward in frequency very slightly. Peak-shifting occurs when the spectral mass surrounding a peak is distributed unevenly. As the ZCPA is a concerted effort to locate peaks accurately, it would be regretful if imprecision in the peak tracker were to nullify this benefit. The course of action taken here is to find peaks in the smoothed ZCPA and then relate them to the maximum in a small neighbourhood of the unsmoothed ZCPA; thus, peaks 7 and 8 would be mapped back to peaks 2 and 4, respectively. Such a scheme exploits spatial averaging to remove spurious peaks and detect broad peaks, but does not sacrifice precision.

Temporal integration can also enforce genuine peaks, provided that the signals in question vary reasonably slowly in time. The peak detector used in this chapter operates on the output of a leaky integrator, I, which is governed by the difference equation

$$a\,\mathrm{I}[k,t] = \mathrm{I}[k,t-1] + \mathrm{ZCPA}[k,t]. \tag{6.15}$$

The leak rate is controlled by the parameter a, which is set to 0.5 in this work and corresponds to a time constant of approximately 100 ms, if frames are recorded 16 times a second. Using larger values for a increases the integration window, and *vice versa*. If the temporal leaky integration is followed by spatial smoothing with a difference of Gaussians, the entire operation is equivalent to a two-dimensional convolution with the kernel shown in Figure 6.10C.

Peak Tracking

The output of the peak detection module, specifically, a set of peak frequencies and their respective values, forms the input to the tracker module, which uses the peaks to extend existing tracks or instantiate new ones. It is the track-formation aspect of McAulay and Quatieri's algorithm to which we most closely adhere in this work. The

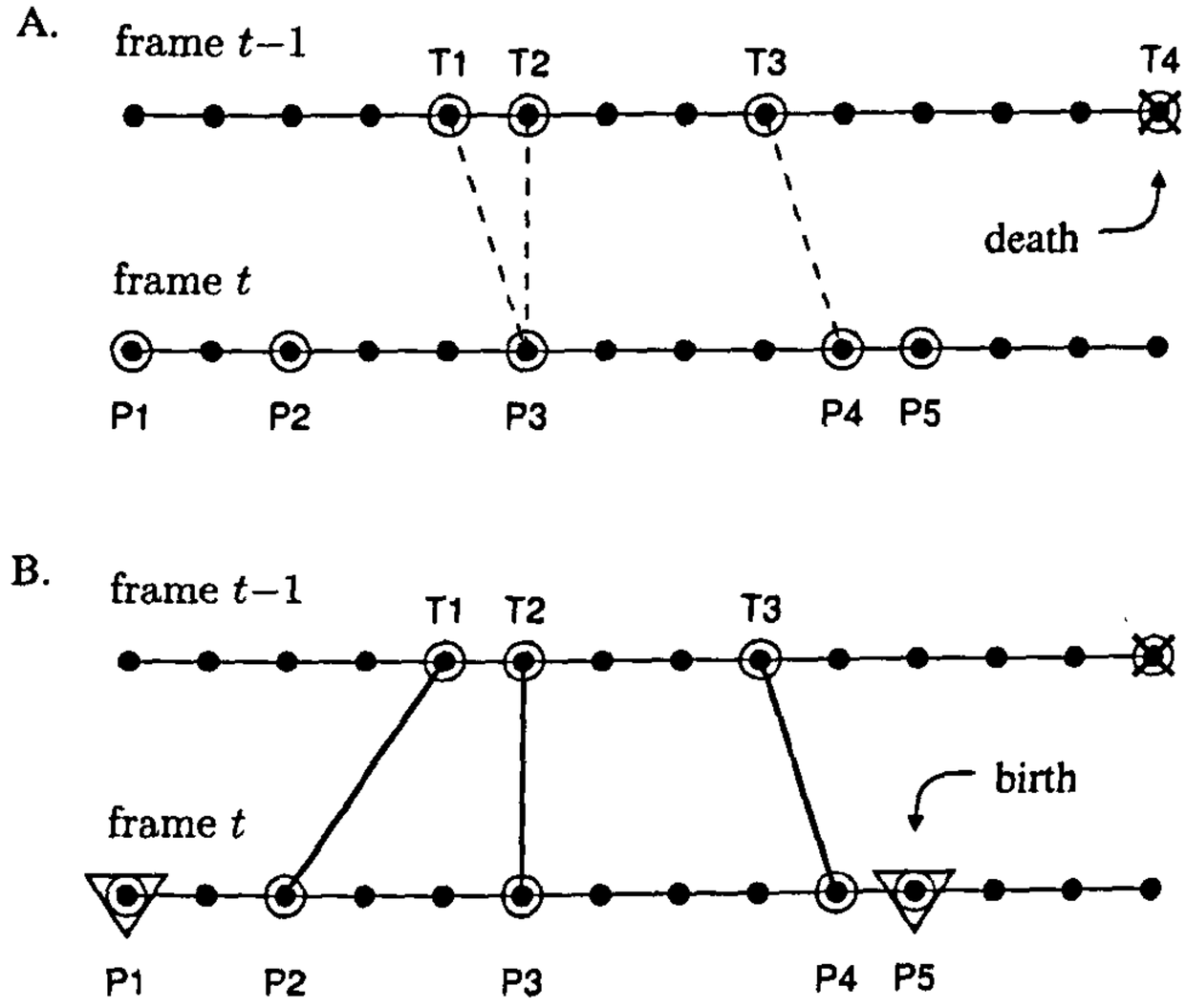


Figure 6.11: McAulay and Quatieri (1986) track formation. Dots show possible peak locations; open circles show locations of peaks; dashed lines show candidate matches; solid lines show definitive matches; inverted triangles mark the first peak of a new track; crosses mark the final peak of a dead track.

routine consists of three steps, which were sketched above, but will now be described in more detail.[1]

In the first step, every live track is provisionally matched against the nearest peak to fall within a *matching interval*, which typically extends a short distance either side of the track's end-point. Figure 6.11A shows how tracks form when the matching interval spans two bins either side: Tracks T1 and T2 form candidate matches with peak P3; track T3 forms a candidate match with peak P4, and track T4 dies, as it is unable to match to any peak. This mechanism tolerates frequency modulation in tracks below a certain depth (measured in bins per frame, or Hertz per second) and vets any unconnected tracks.

In the second step, tracks which are competing to connect to the same peak are resolved. The candidate matches with tracks in the previous frame are examined for every peak, and a *definitive match* is made with the nearest of these. Once a track and peak are definitively matched, both are removed from active consideration. Steps 1 and 2 are repeated until no further matches are possible. Returning to the example in Figure 6.11, a definitive match is made between T2 and P3, and T3 and P4. Step 1 must

[1] For a more formal account of the algorithm, see McAulay and Quatieri (1986).

ZCPA PEAK TRACKER PARAMETERS

Parameter	**Value**
time-frequency convolution	[see Figure 6.10C]
birth threshold	10
continuity threshold	25
matching interval	3 bins
minimum track length	10 frames

Table 6.1: ZCPA peak tracker parameter set.

be repeated to find an alternative match for T1. In this case, it connects to P2 instead. (If no further matches had been available, it would have been terminated.)

In the third step, which is only reached when all tracks in the preceding frame have either been extended or terminated, there may be peaks remaining that do not form part of a track. Each of these forms the first element in a new track. Figure 6.11B shows the final state of every peak and track in Figure 6.11A after step 3 has completed. Peaks P2, P3 and P4 are used to extend tracks T1, T2 and T3, respectively; peaks P1 and P5 instantiate new tracks; track T4 is discontinued.

Added and Proposed Features

The McAulay-Quatieri algorithm was designed for processing relatively clean speech signals, in which most spectral peaks, no matter how energetic, would be identified with the harmonics of voiced speech. However, many of the peaks in a sonar recording arise from a continuous noise background, and we have seen that a threshold is required to reject all but the most prominent. If this threshold is too low, many spurious tracks are born; if it is too high, genuine tracks tend to disintegrate into short strands, and weak signals do not register at all. To overcome this problem, the ZCPA tracker employs two thresholds: the *birth threshold* and the *continuation threshold*. A much higher peak value is required to start a track than to sustain it.

A further feature of the ZCPA peak tracker is the deletion of dead tracks that are considered too short-lived to have arisen from a genuine signal. This measure helps to prevent the time-frequency display and computer memory becoming cluttered with tracks that arise from momentary peaks in the noise floor. However, the removal of short tracks is not always beneficial: sometimes, to a human viewer, a stream of fragmentary tracks is clear evidence of a weak tonal. One possible remedy, which has not been pursued here, is to design a higher-level process, which only disposes of short tracks if they cannot be shown to belong to a larger context.

Table 6.1 summarises the set of parameters that govern the behaviour of the ZCPA peak tracker. Suggested values for each parameter are also listed. These options were used to produce the tracks in Figures 6.9B and 6.9C.

6.3 Timing-based Fine Structure Estimation

A peak tracking algorithm that employs time-frequency continuity as a constraint provides a helpful means of highlighting candidate tonal structure in the surface of the ZCPA. The fine structure of these tonals, which might be useful for deciding how to group them together—for example, very shallow and slow-varying modulations in frequency—cannot be measured in the ZCPA, as the frequency histogram fails to preserve this fine detail. However, the information needed for a finer reconstruction still resides in the strata below the histogram, and we may descend into these as far as necessary to recover it, starting with the circular buffers which feed the histogram, then the output of the zero crossing (and peak) detectors, then the narrowband signals, and ultimately the raw waveform.

We will return to the idea of "ZCPA strata" in the final chapter. For now, it is sufficient to imagine that an operator selects a coarse track in the ZCPA, and an unspecified mechanism locates the relevant detail in one or more layers beneath. Methods for constructing a phase or frequency modulation track from zero crossings divide naturally into two categories: *model-based* and *data-driven*. A model-based approach proceeds along the same lines as the detection routines in the previous two chapters: we devise a model of the channel, possible signals, noise background, etc., and attempt to make the statistically optimal choice with respect to some cost function. A data-driven solution makes few assumptions about the process that produced the data and employs a more generic approach, e.g., a curve-fitting routine.

Two timing-based approaches to frequency tracking are investigated in this section. The first is model-based and draws on the material in Chapters 4 and 5. The second is data-driven and attempts to fit more closely the framework of the ZCPA, as presented in Chapter 3 and in the sections immediately above. These methods will be applied to artificial, rather than recorded, signals. The techniques deemed to be most successful will then be used to analyse recorded signals in Sections 6.4 and 6.5.

6.3.1 Model-based Frequency Tracking

The dominant frequency principle states that "the normalised zero crossing rate is a *weighted average* of the spectral mass" (Kedem, 1986). That is, the zero crossing rate of a Gaussian process locks to the most dominant component in the channel and, for a white noise signal, tends to gravitate towards the band centre frequency. In earlier chapters we encountered the dominant frequency principle as it pertains to zero crossing intervals, where analytical and experimental work showed that the interval distribution reflects the contribution of the signal and noise, weighted according to the signal-to-noise ratio.

Many algorithms extract a component frequency from the zero crossings directly, by halving the reciprocal of the interval between two consecutive crossings, or something similar (Sekhar and Sreenivas, 2005; Kim et al., 1999; Ghitza, 1988). The dominant frequency principle evidently has implications for the estimation of frequency using zero crossing intervals or the phase derivative, especially at low SNRs. We could ask,

for example, "What frequency does a five millisecond interval between two successive zero crossings suggest?" In the naïve scheme, a 5 ms interval would map to a frequency of 100 Hz. However, knowing further that the interval was received in severe noise, in a band-pass channel centred at 110 Hz, we might instead conjecture that the noisy channel "bent" the true frequency towards its centre and adjust our estimate downwards accordingly. The next two sections explore how such an adjustment might be achieved in a principled fashion using results from the preceding chapters.

6.3.2 Maximum Likelihood Frequency Estimation

The interval detectors described earlier chose between two hypotheses on the basis of an observed zero crossing interval: H_0, the interval is due to noise; or ii) H_1, the interval is due to a mixture of signal and noise. In each case, the hypothesis selected was given by

$$\arg\max_j P(H_j \mid i_c)$$

We can of course extend this principle to choose amongst any number of hypotheses. In particular, we can label each possible signal frequency as a hypothesis, e.g., $H_{f_c=100}$, $H_{f_c=101}$, $H_{f_c=102}$, and then the maximum *a posteriori* hypothesis becomes

$$H_{f_c=\hat{f}_c} = \arg\max_f P\left(H_{f_c=f} \mid i_c\right) \tag{6.16}$$

$$= \arg\max_f p_{I_c}\left(i_c \mid H_{f_c=f}\right) P(H_{f_c=f}). \tag{6.17}$$

The essence of statistical estimation is therefore identical to that of detection. In fact, detection can be thought of as an estimation problem, in which the only parameter refers to the signal's presence and can assume the values "yes" or "no".

One possibility at this stage is to assume that all frequencies are equally probable, thus rendering $P(H_{f_c=f})$ constant so that it does not affect the argument of the maximum. This leads to the *maximum likelihood estimate* for frequency,

$$\hat{f}_{c\,ML} = \arg\max_f p_{I_c}\left(i_c \mid H_{f_c=f}\right). \tag{6.18}$$

Maximum likelihood estimation is useful in many instances because it does not refer to the prior distribution of the variable to be estimated. However, the maximum likelihood estimate does not always optimise a continuous parameter in a satisfying way, as we can demonstrate in the present scenario. Suppose the true frequency of a signal is 700 Hz. A maximum likelihood method will maximise the occurrence of an exact estimate of 700 Hz; in that sense it is optimal. However, all incorrect estimates, including, e.g., 699 Hz and 7 kHz, are considered to carry an equal cost. This is clearly an unacceptable criterion for frequency estimation, as we really intend the estimate to be optimal in the sense of being close to the true frequency as often as possible. This is the target of the *Bayes estimate* discussed next.

6.3.3 Bayes Optimum Frequency Estimation

In the most basic formulation of this problem, all the signal properties are known except the tonal frequency, for which a prior distribution is available. The Bayesian frequency estimate minimises the expected value of a *risk function*, given i) an observation or series of observations, and ii) a set of fixed conditions, Θ. In this case, the risk function is chosen to be the squared difference between the estimated frequency and the true frequency. The fixed conditions include the noise power spectral density, signal level and the analysis filter squared magnitude response. Whalen (1971, page 322) describes the procedure for finding the (minimum mean squared error) Bayes estimate for an unknown signal parameter, if all the other parameters are known.

Let $\mathcal{F}$ denote the set of frequencies that it is possible for the component to assume. The objective is to minimise the expected squared error between the true frequency, f_c, and our estimate, $\hat{f}_c$, that is[1]

$$\hat{f}_c = \arg\min_f \left[E\{(f - f_c)^2 \mid i_c, \Theta\}\right] \tag{6.19}$$

$$= \arg\min_f \left[f^2 - 2f \int_{\mathcal{F}} f_c\, p(f_c \mid i_c, \Theta) df_c\right]. \tag{6.20}$$

As the function to be minimised is quadratic and convex with respect to f, it has a single global minimum at

$$\hat{f}_c = \int_{\mathcal{F}} f_c\, p(f_c \mid i_c, \Theta) df_c. \tag{6.21}$$

The distribution for f_c conditioned on i_c has not arisen in this work; nevertheless, we possess the distribution for i_c conditioned on f_c, namely, the interval distribution encountered in Chapter 4, in which f_c is a parameter. Bayes theorem can be used to rewrite (6.21) in terms of the interval distribution and a prior distribution governing frequency, $p(f_c)$. If the prior frequency distribution is uniform, then the Bayes optimum frequency estimate is

$$\hat{f}_c = \frac{\int_{\mathcal{F}} f_c\, p(i_c \mid f_c, \Theta) p(f_c)\, df_c}{p(i_c \mid \Theta)} = \frac{\int_{\mathcal{F}} f_c\, p(i_c \mid f_c, \Theta)\, df_c}{\int_{\mathcal{F}} p(i_c \mid f_c, \Theta)\, df_c}. \tag{6.22}$$

The integral in (6.22) takes the form of a centroid[2], which, in this case, is difficult to solve analytically. The fact that it is a centroid does, however, suggest a numerical approach: populate the *rows* of a matrix with interval probability density functions, $p(i_c \mid f_c, \Theta)$, for a series of f_c uniformly-spaced over the range of $\mathcal{F}$, and evaluate the centroids of the *columns*. The latter is a function that maps an observation, i_c, to the (approximate) Bayes estimate, $\hat{f}_c$, and is referred to as an *adjustment curve*. This is to be contrasted with the naïve mapping,

$$\hat{f}_c^{\text{naive}} = \frac{1}{2i_c}. \tag{6.23}$$

[1] We shall now drop the notation $P(H_{f_c=f})$ in favour of $p(f_c)$.

[2] The centroid is a generalisation of the mean. The centroid of $f(x)$ is $\frac{\int x f(x) dx}{\int f(x) dx}$.

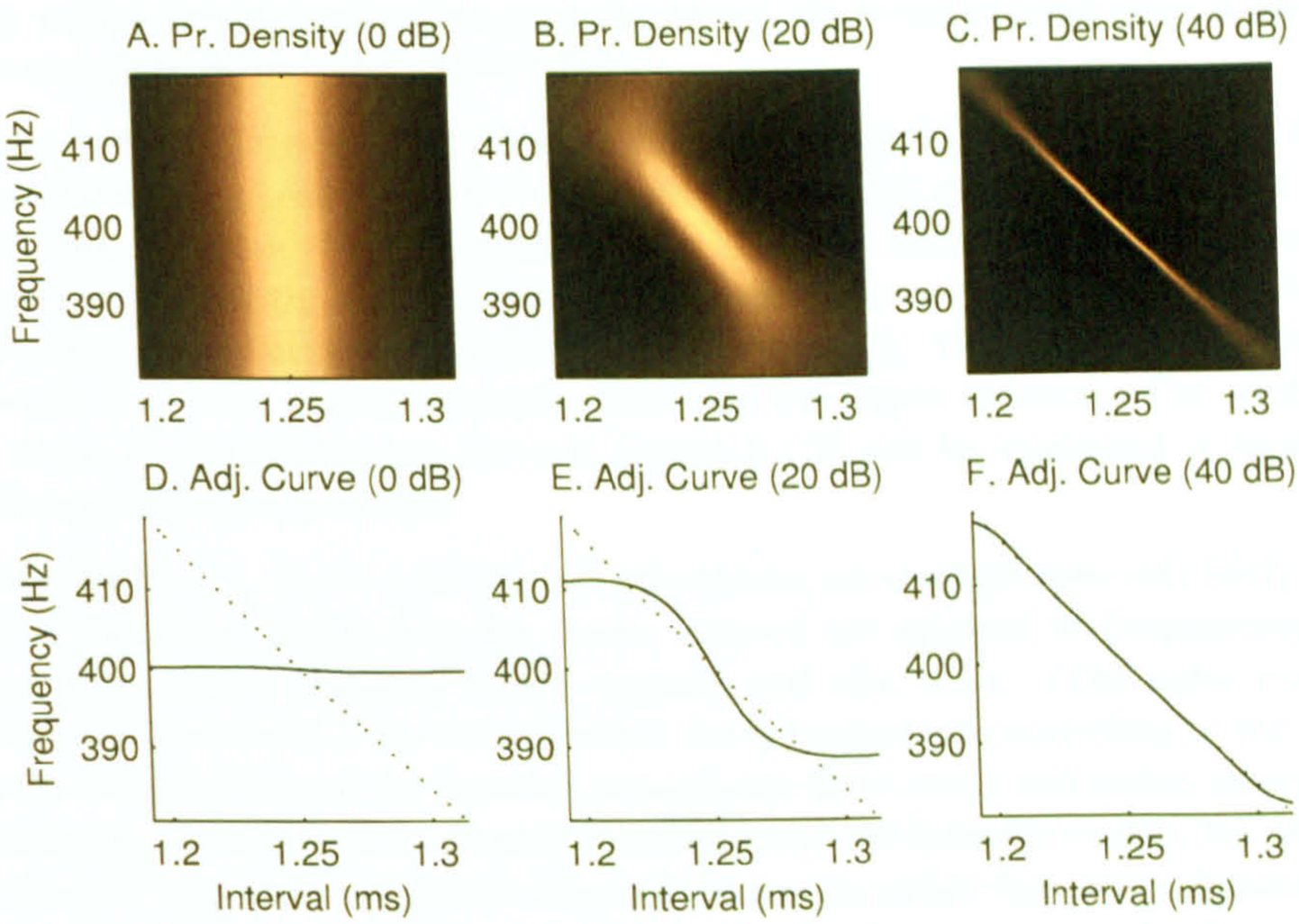


Figure 6.12: A–C) interval-frequency joint probability density: lighter regions show more probable pairings; D–F) interval-to-frequency adjustment curves: Bayes (solid) and naïve (dotted).

Visualising the Adjustment Curves

The images in Figures 6.12A–C show the joint probability density function for tonal frequency (ordinate) and interval duration (abscissa), for a tonal signal received against a white Gaussian noise background, where the frequency is uniformly distributed in the 381 Hz–419 Hz range[1], and the intervals are measured in the output of a narrowband filter. The impulse response of the analysis filter is constructed according to the MGMM in (5.24), which, in the frequency domain, places the the centre frequency at 400 Hz and provides a 3 dB bandwidth of approximately 21 Hz. The narrowband signal-to-noise ratios are intended to reflect, respectively, noise-dominated (0 dB), evenly-mixed (20 dB) and signal-dominated (40 dB) scenarios.

A row of pixels in each image is proportional to the conditional interval probability density function. The density function shown in Figure 6.12A corresponds to a mixture dominated by noise. As the signal exerts very little effect upon the intervals, the interval p.d.f. in each row is the same. Consequently, the centroids taken along the columns always fall at the centre frequency, and the adjustment curve, shown underneath in Figure 6.12D, is (almost) a constant 400 Hz. This is an intuitive result: in the absence

[1] These values were arbitrarily chosen to match the 10 dB bandwidth of the analysis filter. In the context of a bank of filters, the crossing points could be used instead. Alternatively, a separate probabilistic model could be used to predict the signal frequencies.

of any evidence concerning the signal frequency, the mean squared error is minimised by choosing the prior expected frequency.

Figure 6.12B displays the joint density function associated with a moderate SNR. In this scenario, the signal influences the intervals to some extent, and the p.d.f. in each row is spread around the dominant interval. There is little spread in the intervals for frequencies around the channel centre, where the post-analysis SNR is high, but at the filter edges, the intervals are distributed more broadly. Computing the column-wise centroids gives the adjustment curve based on the Bayes estimate. The backward S-like shape of the adjustment curve in Figure 6.12E can be explained in terms of the dominant frequency principle.

At the centre, i.e., about 1.25 ms, the adjustment curve maps intervals fairly directly. Intervals slightly longer than the centre interval are mapped to frequencies slightly *lower* than a naïve mapping would suggest, and *vice versa*. (The naïve mapping is shown as a dotted line.) As noted earlier, the estimator acts according to the principle that the noise has biased the intervals towards the filter centre and makes an appropriate adjustment. However, at frequencies further removed from the centre, the adjustment curve starts to generate estimates *towards* the centre, rather than away. Intervals in this remote domain are more likely to be the result of noise; consequently, the estimator is less certain and thus more inclined to choose frequencies closer to the centre, in order to minimise mean squared error as described above. (Were this curve to be extended in both directions, each extreme would eventually converge to 400 Hz.)

Lastly, the joint probability density function for the 40 dB SNR mixture is plotted in Figure 6.12C. In this scenario, the signal is dominant and the noise affects the intervals only very mildly—a reversal of the situation in Figure 6.12A. This too leads to an intuitive result: in the absence of any noise, the mean squared error is minimised by a naïve mapping. The adjustment curve in Figure 6.12F takes the form of a reciprocal function everywhere except at the very edges, where the signal is attenuated and the noise becomes influential. Even in the low and high SNR cases, the adjustment curves retain some evidence of a sigmoidal shape.

Visualising Adjustments in the Estimates over Time

Adjustment curves such as those in Figures 6.12D–F provide a static impression of how intervals are mapped to frequencies at a particular SNR. One can gain an alternative insight into this process by synthesising a mixture of tone and noise, passing the mixture through an analysis filter, and recording two time series: the intervals of the process mapped to frequencies by i) a naïve approach and ii) a Bayesian approach. Figure 6.13B plots the frequency estimates for one-second of a synthetic 402 Hz tone mixed with Gaussian noise at 20 dB narrowband SNR. We identify the time of an interval by its centre.

Naturally, most of the comments that apply to the two frequency tracks in Figure 6.13B we already made in connection with the adjustment curve in Figure 6.12E. The series always intersect around 400 Hz where the naïve and Bayes estimates coincide, and for small variations around the centre, the Bayes estimates tend to "exaggerate"

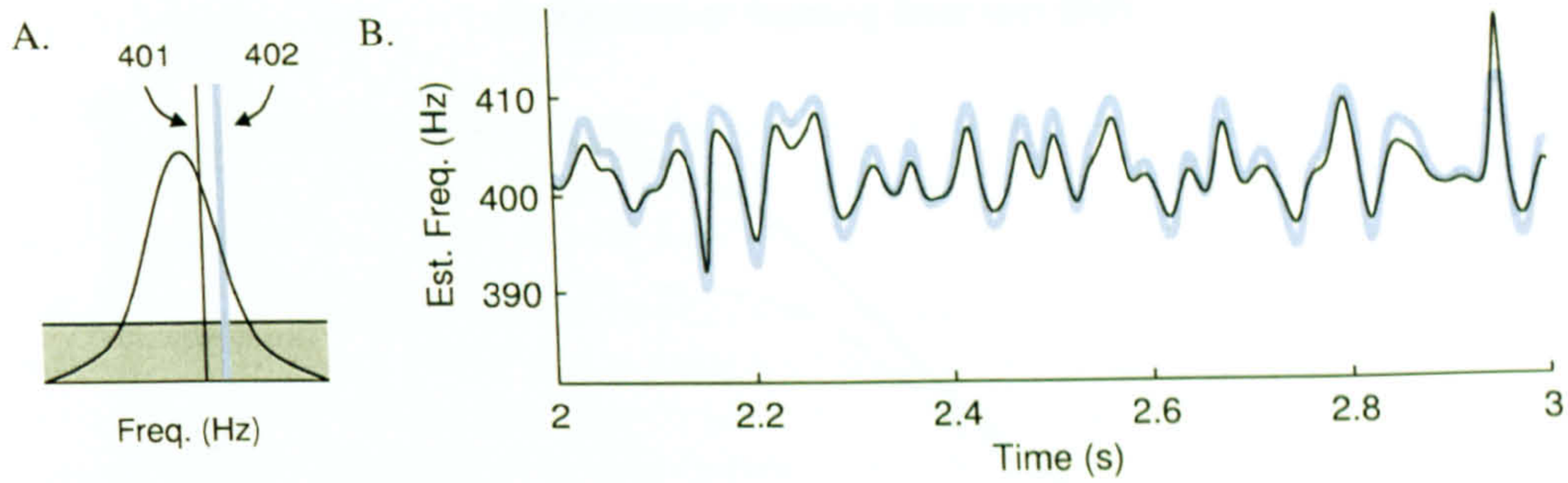


Figure 6.13: A) (*not to scale*) noise in the filter causes the true signal component (light blue, thick) to be measured nearer the filter centre (black, thin); B) in an effort to undo this effect, the naïve time-varying frequency track (black, thin) can be adjusted to account for the noise floor, giving the Bayes track (light blue, thick).

the directly-measured frequencies to counteract the effect of the noise floor. For outlying measurements, such as the peak around 2.9 seconds, the Bayesian method underestimates the displacement of the component from the centre in order to minimise mean squared error.

Experimental Results and Analysis

The first experiment measures the mean squared error between the estimated frequency and the true frequency as a function of narrowband SNR. The analysis filter is centred on 400 Hz and the sinusoid frequency is chosen according to a uniform distribution spanning the 10 dB bandwidth of the filter, namely, 380.68 Hz–419.32 Hz. In both the naïve and Bayes schemes, the intervals exiting the analysis filter are gated into this bandwidth prior to estimation to prevent large spikes affecting the mean squared error. The results of this experiment are plotted in Figure 6.14A. From this graph it is clear that the Bayes estimate consistently matches or improves upon the naïve estimate at all signal-to-noise ratios.

The second experiment examines more closely the particular conditions under which the Bayes estimator outperforms the naive estimator. This task involves holding the signal-to-noise ratio fixed and measuring the squared error at particular frequencies, as opposed to averaging the error over all frequencies. Figure 6.14B shows the mean squared error when the pre-filter SNR is 0 dB. As the post-filter SNR is much lower, the naïve estimator is essentially measuring (gated) random intervals and mapping them directly to frequencies. The error is lower when the true frequency is nearer 400 Hz, simply by virtue of the fact that the random intervals are distributed around this value—the signal has little effect. Informally speaking, the Bayes estimator has already "realised" this, and for that reason, always chooses 400 Hz. (We shall return to this issue shortly.)

The results obtained for the 20 dB and 40 dB SNR scenarios are of greater interest. In the 20 dB case, the squared error in the Bayes estimate is smaller than that of the naïve estimate when the signal is placed away from the band centre, but the converse holds for

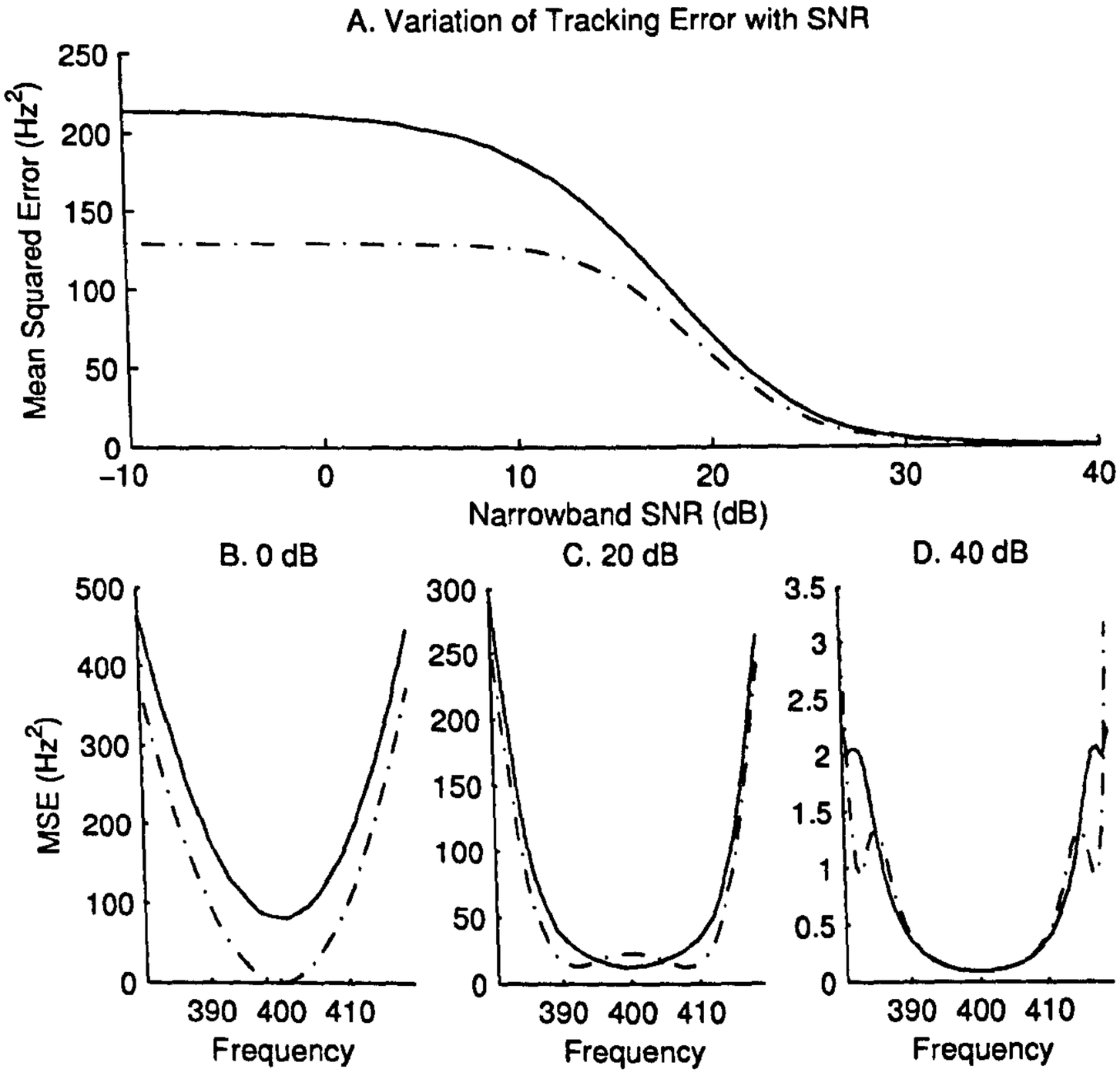


Figure 6.14: Estimation of a tonal frequency, when tonal is placed uniformly at random in the bandwidth of an analysis filter and the SNR is fixed and known (see text). A) the mean squared error between the true and estimated frequencies for the naïve (solid) and Bayes (dashed) estimators; B–D) break-down of the mean squared error in terms of frequency at three SNRs.

signals near the band centre. The reason for these differences can be explained in terms of the dominant frequency principle: the presence of white noise in the filter causing intervals to gravitate towards the band centre. Consider first the situation in which the signal is near the centre. The naïve estimator will (rightly) convert the interval to a frequency in direct fashion, but the Bayes estimator will be inclined to view the interval as having been corrupted by noise, and (wrongly) adjust it away from the centre. Consider now the situation in which the signal has been placed away from the band centre, at either a lower or higher frequency. This time, the naïve estimate fails to adjust for noise floor, incurring a large penalty in squared error, whereas the Bayes estimate accounts for the effect of the noise upon the intervals by means of the adjustment curve.

Critical Discussion

In closing, several objections may be raised against Bayesian frequency estimation based on zero crossing intervals, despite the fact that the results in Figure 6.14A suggest an improvement over the naïve estimate.

First, to compute a Bayes estimate effectively requires a comprehensive description of the signal and noise model, and a prior distribution over frequency. It is unlikely that such a model would be available in practice, although one could estimate the latent variables of the model in an online fashion[1]. A considerable number of computations are needed to compute an adjustment curve.

Second, at low SNRs, the Bayes estimate achieves a large part of its gain in performance by choosing the expected frequency *a priori*, whereas at high SNRs, there is only a negligible difference in performance. This implies that there is only a narrow range of signal-to-noise ratios over which Bayes estimation can be considered worthwhile. It is questionable whether an estimator which returns a 400 Hz frequency track in severe noise conditions simply to minimise an average cost function is actually very helpful, even if it is technically optimal.

Third, a notable feature of Figures 6.14C and 6.14D is that most of the performance gain is associated with signal frequencies in the band edges. In the context of tracking in the ZCPA, given that the analysis filters exhibit a certain degree of overlap, it seems likely that a dominant component would be tracked near the centre of a proximate filter, rather than in the edge of a remote filter. Furthermore, a ZCPA parameterisation suitable for sonar analysis employs a DFT with a much narrower bandwidth than the filter used in this example. In narrowband channels, the dominant frequency principle exerts very little influence over the intervals.

6.3.4 Interpolating Intervals with a Cubic Spline

The data-driven approach that we shall consider next is based on *spline interpolation*. The circular buffers of the ZCPA, the contents of which are periodically used to produce the histogram, hold the most recent intervals and peaks recorded in each channel. The size of each buffer is equal to the interval-peak window (*cf.* §3.3.3). This layer of the ZCPA thus stores potential data points through which to reconstruct a frequency track. This fine structure estimation is a two-part process: first data points are computed, then a polynomial function is fitted to them.

Retrieving Data Points from the Circular Buffers

At every time step in which the ZCPA is recorded, there resides in each channel buffer a certain population of intervals and peaks. Let us assume that at a particular time step a coarse peak track is live and, furthermore, there is a mechanism for associating the high-level track with a low-level buffer. Several ways to form a data point, d, from the

[1] Refer to Whalen's discussion of *generalized likelihood ratio detection* (Whalen, 1971, page 352).

intervals in the circular buffer suggest themselves. These we may straight-forwardly list, using $i_c[l]$ and $e[l]$ to denote respectively the l-th interval and peak in a buffer of length L:

1. the mean interval,

$$d = \frac{1}{L}\sum_{l=1}^{L} i_c[l];$$

2. the peak-weighted mean interval,

$$d = \frac{\sum_{l=1}^{L} e[l]\, i_c[l]}{\sum_{l=1}^{L} e[l]};$$

3. the harmonic mean interval,

$$d = \left[\sum_{l=1}^{L} \frac{1}{i_c[l]}\right]^{-1};$$

4. the median interval;

5. or any of summary statistics 1–4, formed from intervals that have been adjusted to account for the dominant frequency principle, using the material set out above.

Any of these five options are a possibility (although the fifth demands a rather exact knowledge of the signal and noise statistics), and many more are conceivable. As we are primarily interested in temporal processing, and there is not space for a detailed appraisal of each alternative, we shall elect the simplest of these: the mean interval.

Fitting a Polynomial Curve

Having obtained a series of interval measurements for each frame of the ZCPA, there remains the problem of finding a smooth function which passes through or near each data point, and provides a frequency estimate at all times between the first and last data point. One course of action is to fit an nth-order polynomial curve through all the data points, leading to a function of the form

$$\hat{i}_c(t) = a_n t^n + a_{n-1} t^{n-1} + \ldots + a_n t + a_0, \tag{6.24}$$

in which $a_0, \ldots, a_n$ are optimised coefficients[1], and there is a corresponding naïve continuous-time frequency track,

$$\hat{f}_c(t) = \frac{1}{a_n t^n + a_{n-1} t^{n-1} + \ldots + a_n t + a_0}. \tag{6.25}$$

[1] This can be carried out quite easily, using the least mean squares approach described in Section 5.2.4. Rather than finding a linear combination of Rayleigh density functions to minimise a cost function, we find a linear combination of monomials in t, i.e., 1, t, t^2, and so on. This kind of polynomial curve fitting is ubiquitous; for a particular application to zero crossing intervals, see Sekhar and Sreenivas (2005).

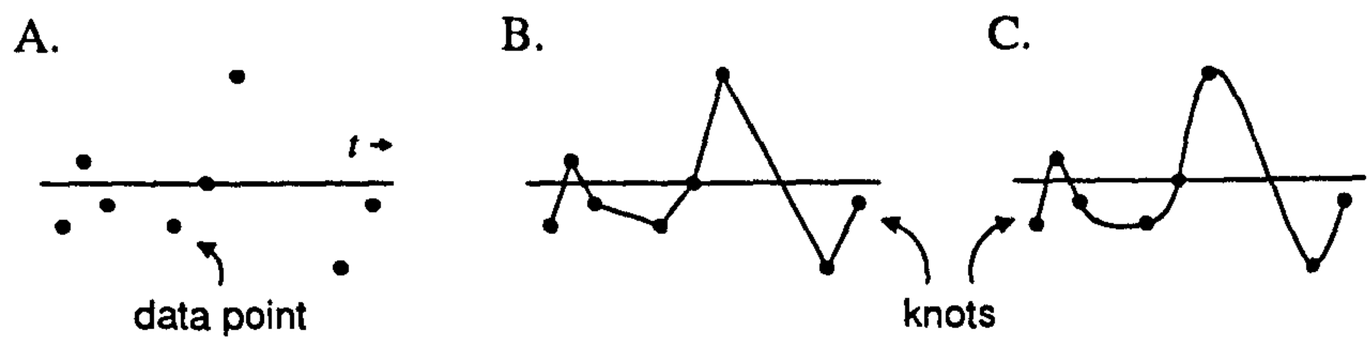


Figure 6.15: Spline interpolation. A) data points; B) linear spline; C) cubic spline.

Alternatively, one could fit the polynomial through the reciprocals of the data points to obtain a track for $\hat{f}_c(t)$ directly. Note that this track would *not* be identical to that obtained using (6.25), but both are admissible solutions.

There is good justification to reject a polynomial fitting of a fine frequency track. The first problem arises when an otherwise-smooth series of data points is interrupted by one or more outlying values, e.g., shot noise. Least mean squares approaches are based on minimising an average cost, so a sample displaced from the smooth curve by a large magnitude severely affects the fit. As outlying values can be removed in advance, for instance, by a median filter or threshold, this problem is not insurmountable. (The strategy of removing untrustworthy data points from the series is considered in the next section.)

A second, more serious problem concerns *model order selection*, that is, the choice of whether the function to fit to the data points is to be linear, quadratic, cubic, and so on. If the model order is too low, then the curve will be too inflexible to represent the data; on the other hand, if the model order is too high, the curve might follow random fluctuations in the data. Selecting a model order on the basis of the data points alone, without any knowledge of the underlying, real-world process that produced them, is an advanced problem beyond the scope of this work.

Fitting Cubic Splines: Background

Spline interpolation fits a string of low-order polynomials (placed end-to-end) to the data points of a series, which in this context are referred to as *knots*. (Contrast this with the approach described above, which attempts to fit one, high-order polynomial to the entire series.) A *linear spline* interpolation, for instance, interpolates a linear segment between successive pairs of samples; the entire series is therefore a piecewise linear function defined everywhere between the first and last knot. Figure 6.15B illustrates a linear spline.

A *cubic spline* is extends this principle to generate a smooth[1] curve: between every pair of consecutive samples a cubic function of time is interpolated. There are only two equations to secure each cubic polynomial, i.e., the values at the knots; so, initially, each polynomial is underdetermined. This underdetermination is solved by insisting

[1] The interpolant appears smooth in a visual sense. In a technical sense, the first and second derivatives of a cubic spline are continuous, but generally, the third derivative is not.

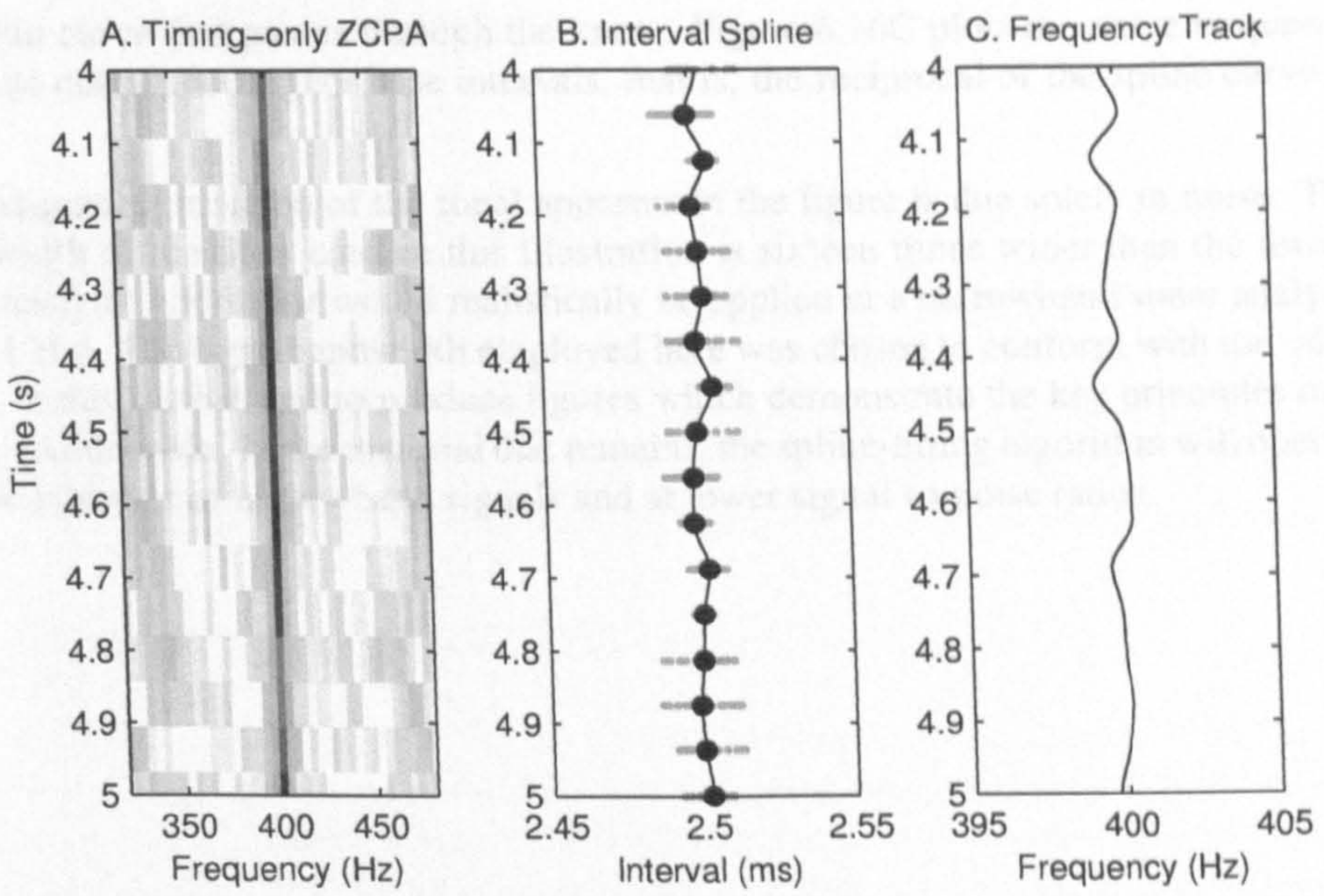


Figure 6.16: Cubic spline interpolation through interval data points. A) appearance of tonal in the ZCPA; B) twenty intervals recorded at each time step (grey dots), mean interval data points (black dots), and a cubic spline through the data points (solid line); C) naïve frequency track (reciprocal of spline). Note the various units and ranges used on the abscissae.

that the first and second derivatives at the end points of successive segments match, as well as including a constraint that states that the second derivatives at the first and last knot are zero. This formulation results in the *natural cubic spline*, an example of which is sketched in Figure 6.15C.

The procedure for determining the coefficients of a cubic spline, and its theoretical justification, are described in numerous texts and need not detain us here (Ayyub and McCuen, 1995; Press et al., 1992). In MATLAB spline interpolation is carried out by the `interp1` or `spline` functions.

Fitting Cubic Splines: Example

Figure 6.16A shows how a tonal appears in the timing-only ZCPA over a one-second duration, when i) the tonal is centred at 400 Hz in white Gaussian noise, ii) the narrowband SNR is 36 dB, and iii) the parameters of the analysis filter within the ZCPA are configured to match that used in the Bayes estimation study above. The DFT analysis cell in which the signal is principally resolved is 16 Hz wide and centred on 400 Hz.

In Figure 6.16B, the intervals extracted at each time step of the ZCPA are plotted as twenty grey dots, and the mean interval is plotted as a larger, black dot. The mean interval dots form the knots of a natural cubic spline, which is drawn on the figure

as a thin curve that passes through the knots. Figure 6.16C plots the naïve frequency estimate corresponding to these intervals, that is, the reciprocal of the spline curve in (B).

The frequency variation of the tonal apparent in the figure is due solely to noise. The bandwidth of the filter used in this illustration is sixteen times wider than the lowest DFT analysis width that would realistically be applied in a narrowband sonar analysis (i.e., 1 Hz). The large bandwidth employed here was chosen to conform with the other work in this section and to produce figures which demonstrate the key principles on a more visible scale. In the material that remains, the spline-fitting algorithm will operate on the intervals of narrowband signals and at lower signal-to-noise ratios.

6.4 Repairing Fine Structure Tracks through Transients

The spline interpolation scheme proposed above involves connected a series of knots with low-order polynomials to form a twice-differentiable continuous track. If the instantaneous frequency track is disturbed momentarily, for instance, by a transient event or a drop in SNR (e.g., due to envelope fluctuation), then we can consider discarding unreliable knots to produce a *non-uniform spline*. The goal of this section is to develop a transient detector to automate the process of removing potentially damaged knots, and then to demonstrate this algorithm working in realistic noise conditions.

A Cautious Analogy from Auditory Scene Analysis

Before designing an algorithm, it is appropriate to recall from Chapter 2 the Gestalt principle of *closure*. Closure refers to the perceptual completion of a form that has been obscured by another object. The particular instance of closure which is relevant to the present discussion is *tonal completion*—the perceptual restoration of harmonics that have been interrupted by noise. For example, speech is perceived as a continuous stream when interrupted by noise bursts. Auditory scene analysis "interpolates" the missing harmonic segments (Bregman, 1990), as Figure 6.17A illustrates. There is also evidence that the tone entering and exiting the obscured region must form a smooth, continuous track to be reconstructed (Ciocca and Bregman, 1987).

There are, however, important differences between the tonal interpolation procedures proposed next and the psychophysical effect of closure. Gestalt closure only occurs where there is evidence that a tonal has been *obscured*; silent interruptions in a tone are not restored (Figure 6.17B). The algorithm in this section does not decide where tonal segments occur, nor how to connect them together, nor whether good continuity is maintained. (If anything, the ZCPA peak tracker and spline interpolation block would

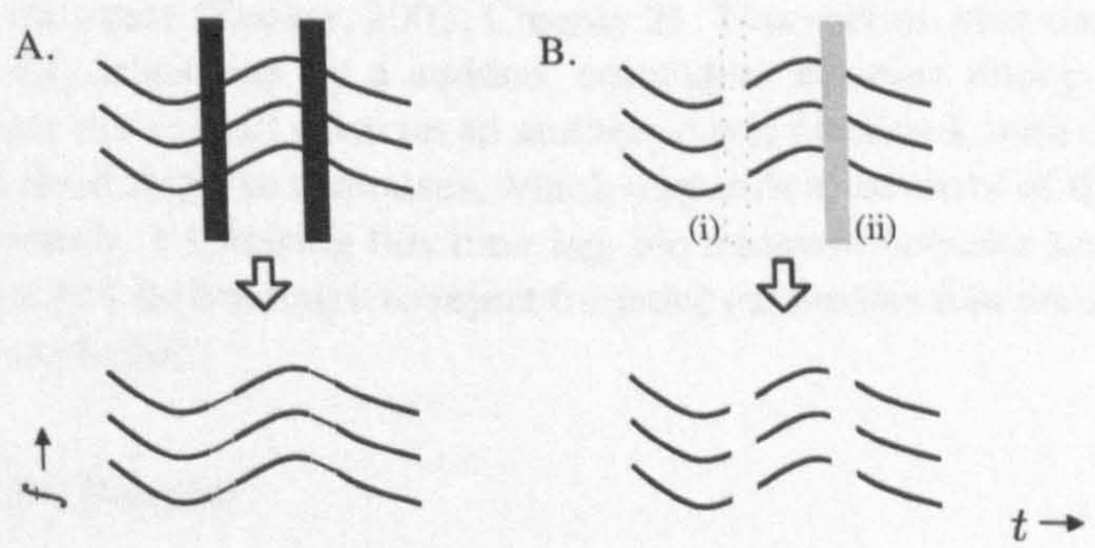


Figure 6.17: A) completion of three tones through two noise bursts—what Miller and Licklider (1950) refer to figuratively as the "picket fence effect"; B) tonals interrupted by (i) silence or (ii) a weak noise burst are not completed.

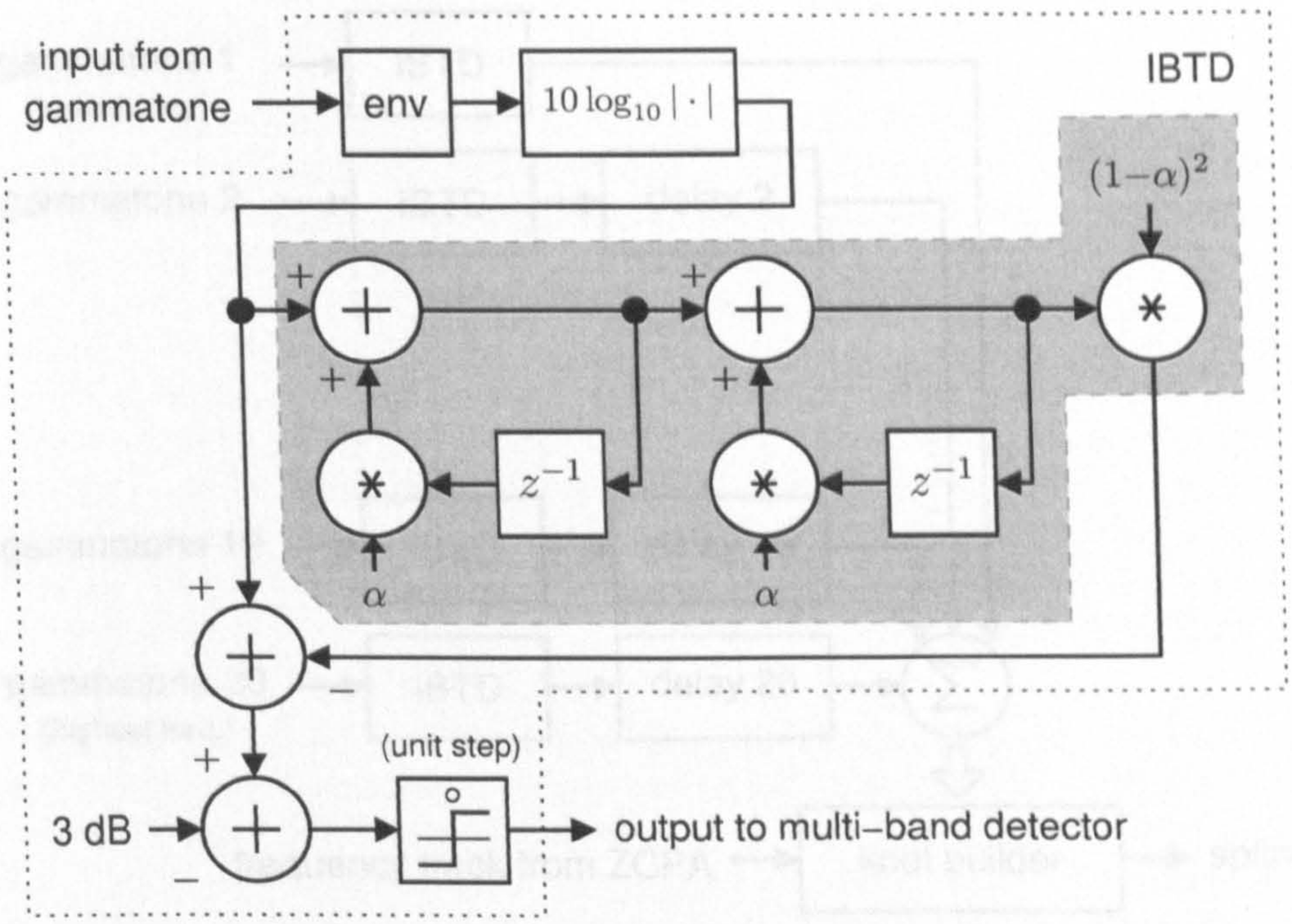


Figure 6.18: Block diagram for the in-band transient detector (IBTD). The grey region is an IIR filter designed to monitor the recent average level.

be responsible for this.) It is simply intended to fill in the corrupted portions of a tonal frequency track during transient events. The analogy is therefore a partial one.

6.4.1 A Rudimentary Transient Detector

A number of auditory-motivated sonar transient detectors and classifiers have been proposed in the literature (Tucker, 2003, Chapter 2). This section proposes a very basic detector to identify transients by a sudden, coincident increase energy across many channels. The detector is built upon on an auditory-style filterbank with relatively wide bandwidths and short impulse responses, which responds in advance of the narrowband surveillance channels. Exploiting this time lag, the transient detector sends a signal to the frequency tracker, instructing it to reject frequency estimates that are about to arrive, as they may be corrupted.

In-band Transient Detector

The front end of the transient detector is a bank of twenty gammatone filters, spaced evenly on an ERB scale between 32 Hz and 4096 Hz. The output of each filter is supplied to an in-band transient detector (IBTD), a schematic of which is provided in Figure 6.18. The operation the IBTD is best described sequentially, starting with the gammatone input at the top left-hand corner of the diagram.

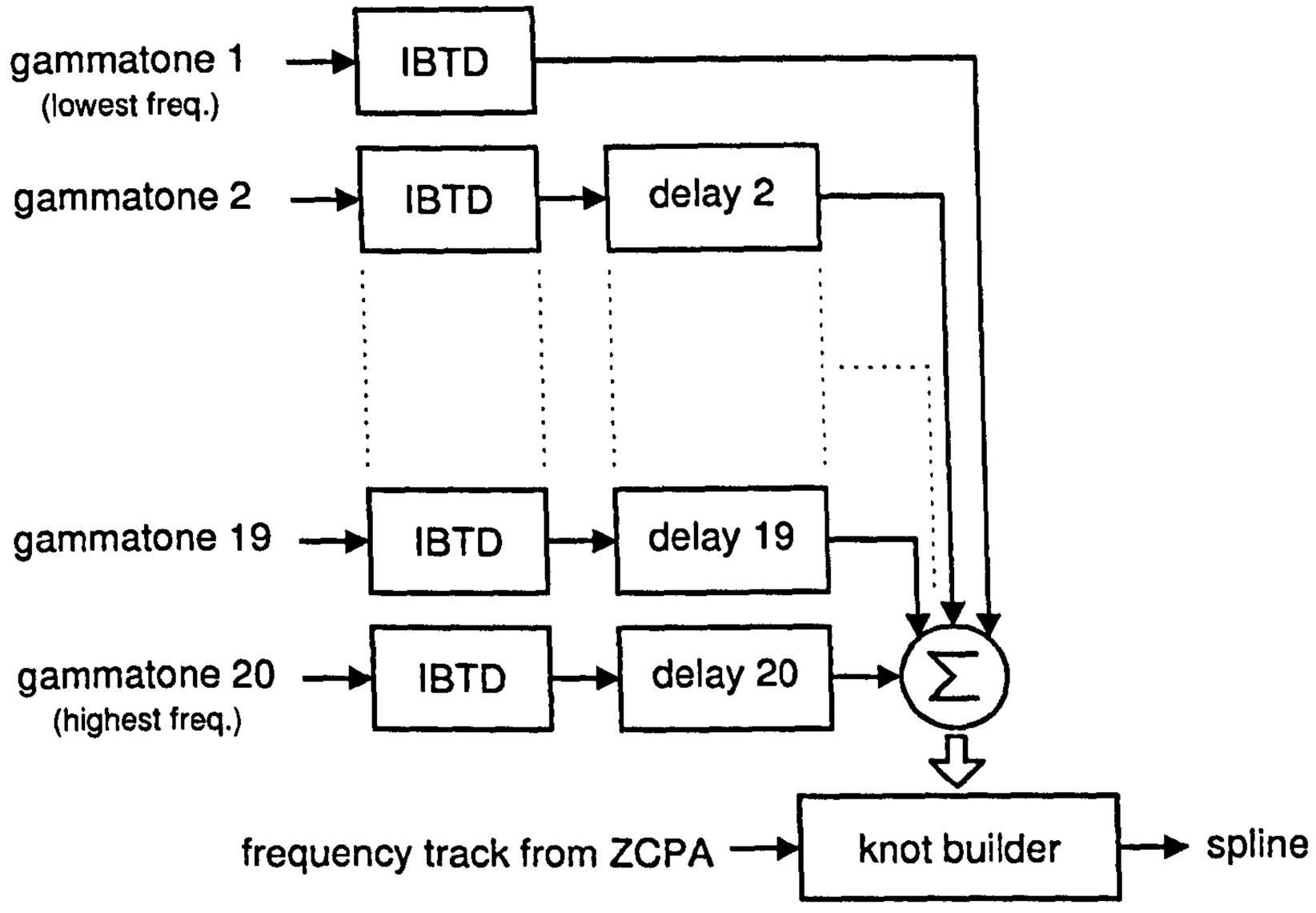


Figure 6.19: Block diagram for the multi-band transient detector. Note that the pathway exiting the lowest-frequency IBTD does not contain a time delay.

The first block computes the envelope of the band-pass signal; the second converts its value to a decibel level. The purpose of the remaining blocks is detect jumps in the log-envelope of the signal that may be indicative of a transient. For instance, an excess of 3 dB above the steady level is equivalent to the linear envelope doubling. The mean signal level is frequency and time dependent, and varies in an unpredictable manner, making it necessary to establish a moving baseline in each channel. The baseline in an IBTD is continually re-estimated by a cascade of two simple one-pole IIR filters, which tracks the mean level in a trailing window. (These components are marked by a grey box in Figure 6.18.) The length of the averaging window is controlled by the α parameter and, in this model, is configured to be in direct proportion with the length of the channel impulse response. At 1 kHz, $\alpha = 0.99$.

The final stage of the IBTD is a hard clip that outputs the extent to which the difference between the rapidly-varying log-envelope and the slow-varying dynamic threshold exceeds 3 dB, or zero if the difference is less than 3 dB. Equivalently, the IBTD does not respond when the linear envelope drops, or jumps by a factor less than two. The detector is therefore sensitive neither to natural undulations in the envelope due to noise nor to low-frequency amplitude modulations; only sudden changes in level, which the adaptive threshold cannot absorb rapidly enough, generate output.

Multi-band Transient Detector

The peak impulse response in each channel of the gammatone filterbank occurs later at lower frequencies, so an ideal wideband impulse will elicit a stream of responses from the IBTDs, which begins at the high-frequency channels and progresses down to the low-frequency channels. Given this lack of synchrony, it is evident that a *static sum* across channels will fail to form a global peak response, as the contributions arrive at different times. The output of the multi-band transient detector is therefore a *delayed sum* of the outputs taken across the bank of in-band transient detectors. The need to envelope-align the filterbank can be addressed in two similar ways. The first is to employ a set of non-causal gammatone filters, the peak responses of which have been aligned to time zero by a introducing an appropriate negative delay (i.e. a lead) (Brown, 1992; Patterson et al., 1988). From a design perspective, the only situations that demand a non-causal approach are those in which the filterbank must respond at *precisely* the moment of the impulse.

The present application does not require an instant response from the filterbank, only that: i) all the gammatones respond in unison to a transient, and ii) they do so far enough in advance of the spline builder to suspend it in real time during an interruption. This leads to the second option: delaying all the IBTD outputs in order that the peaks coincide with the lowest-frequency filter, as Figure 6.19 shows. The gammatone impulse response is given in (2.1). Let B_s denote the parameter B used in the impulse response of channel s. To align the peak response of all the filters with that of channel one, we must delay channel s by

$$(n-1)\left(\frac{1}{2\pi B_s} - \frac{1}{2\pi B_1}\right) \text{ seconds,} \tag{6.26}$$

where n is the filter order.

In the current model, the lowest-frequency IBTD has a peak delay of about 107 ms. Theoretically, a fine frequency estimate in a frame arriving from a ZCPA with 1 Hz DFT bins is delayed by about 500 ms, as the impulse response is one-second long and the peak of the Gaussian window occurs at its centre, although frequency estimates may be affected wherever the analysis window and the transient overlap.

6.4.2 Proof of Concept

To demonstrate the principle of tonal repair, a synthetic 401 Hz tone has been added to a single hydrophone recording containing a transient knock. The mixed signal is plotted in Figures 6.20A and 6.20B, in the time and frequency domains, respectively. The signal is two seconds in duration: the first second is included to ensure that any ringing from the analysis filters—either in the transient detector or the ZCPA—has subsided; only the remaining second, which contains the transient event at about 1.2 s, is shown in the figure.

Figure 6.20C displays the response of all twenty in band transient detectors over the 1–2 s period. The channel centre frequency for each line is measured on the ordinate in ERBs; the height of each bump above this line is measured in decibels, and is scaled

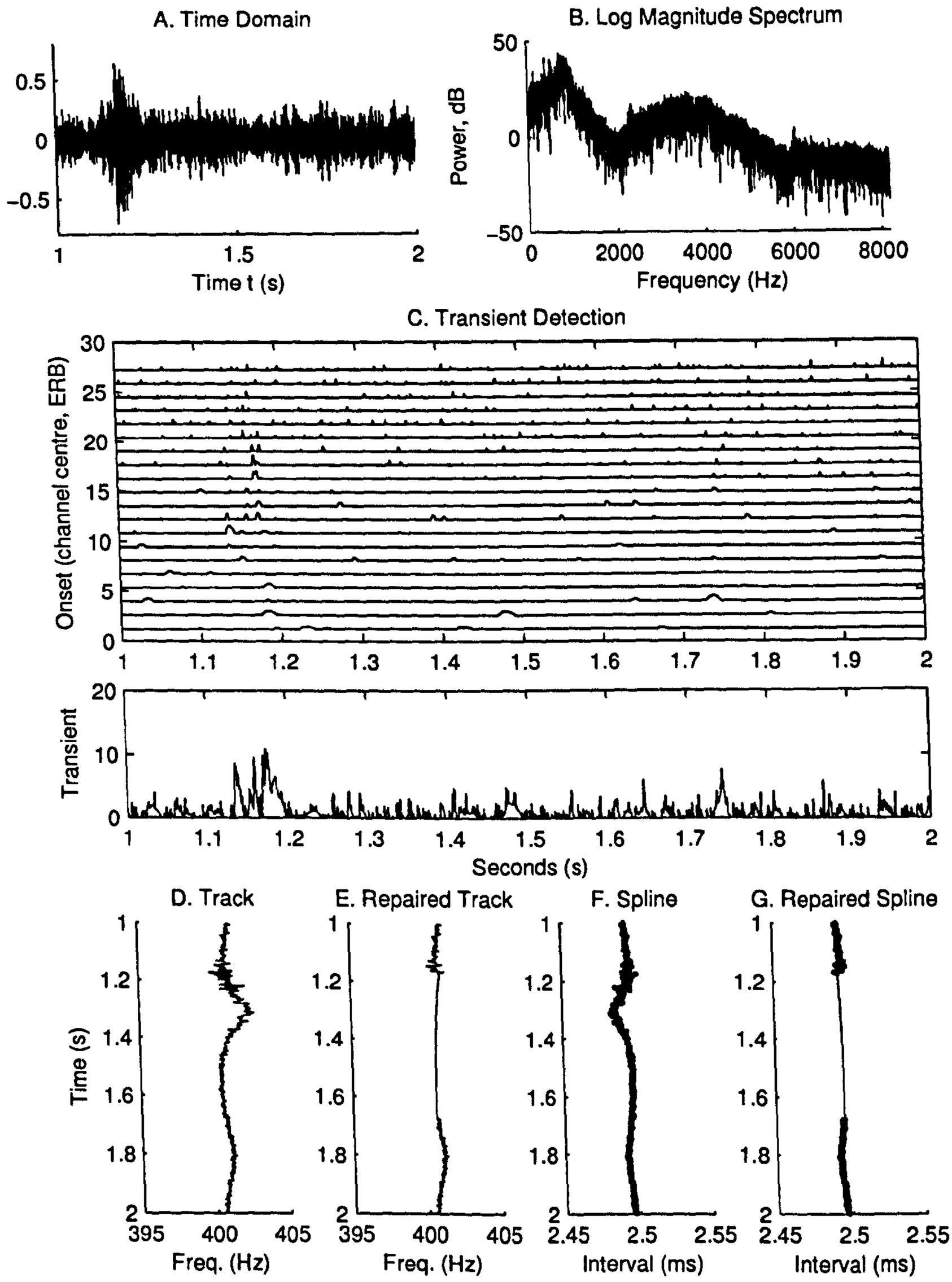


Figure 6.20: A) signal in the time domain; B) signal in the frequency domain; C) response of twenty in-band transient detectors with the summary response beneath; D) damaged track; E) repaired track; F) spline with all knots; G) spline with damaged knots removed.

so that the space between two lines corresponds to 5 dB. The abscissa measures time, and the signal in each band has been shifted to account for any artificial delay. All the channels intermittently generate spikes[1], but a significant volley of synchronised spikes occurs between 1.1 s and 1.2 s, during the knock. These coincident spikes contribute to a peak in the summary plot, shown in the lower portion of Figure 6.20C, which is chiefly concentrated around 1.17 s. We shall assume that the multi-band detector generates a "transient warning signal", whenever the summary output exceeds ten[2].

The destructive effect of the transient on the frequency track is apparent in Figure 6.20D, which is, in turn, based on the spline drawn in Figure 6.20F. (In these figures, time is marked down the ordinate and frequency along the abscissa, reflecting the organisation of a ZCPA display.) Momentarily, the zero crossing intervals in the ZCPA are those of the transient, not the tonal, and this causes the track to be displaced by as much as 1.5 Hz at 1.3 s. Beyond 1.5 s, the track reverts to a steady 401 Hz, with mild fluctuations due to stationary, additive noise.

The final step in this discussion simply unites the comments set forth in the preceding two paragraphs: the multi-band detector finds a transient at approximately 1.17 s, and the spline is damaged because knots placed during the 1.1 s–1.5 s period are unreliable. The transient warning signal from the detector—corresponding to the wide arrow on the diagram in Figure 6.19—has the effect of suppressing knots for 500 ms, which is equal to half the impulse response duration of the ZCPA analysis. The restored spline is plotted in Figure 6.20G, and this is, in turn, used to construct the repaired frequency track shown in Figure 6.20E.

[1]Here, the term "spikes" is used of the curve and need not refer to nerve action potentials.

[2]A threshold of ten has admittedly been chosen after inspecting the data. It is not *signal-dependent*, however, but rather depends on the time constants and thresholds in the IBTDs, and the number of channels.

6.5 Grouping Fine Structure Tracks

Once a set of tonal fine frequency tracks has been extracted (and restored), we may search for subsets that exhibit a common modulation pattern, in order to show that they have arisen from a common source or channel. The work presented in this section proceeds in two different directions. The first involves extracting as many frequency tracks from the noisy signal as possible, and then grouping them according to some similarity metric. The second involves extracting just one or two reliable frequency tracks and then actively searching for similar tracks in the noisy signal.

6.5.1 Passive Comparison to Find Similar Tracks

Non-uniform Sampling

Measuring phase variations using zero crossings is rather different to approaches based on the Hilbert transform or DFT, as it relies on *non-uniform sampling* (Sekhar and Sreenivas, 2005). Standard uniform sampling schemes measure how much the phase has changed at fixed points in time. (For an overview of standard techniques, see Cohen (1995).) By contrast, non-uniform sampling schemes measure how much time has elapsed at fixed points in the signal phase (i.e., zero crossings).

The ZCPA examples described above compute each frame from the twenty most recent upward zero crossing intervals. In the time span over which these intervals occur, the signal phase has advanced by 40π, which is equivalent to stating that the signal phase has advanced by 2π across the duration of a mean interval. If the mean interval is 0.01 s, e.g., then the instantaneous frequency is approximately

$$\frac{2\pi}{0.01} = 200\pi, \text{ radians per second, or } 100 \text{ Hz}.$$

Evidently we are still working with a measure of instantaneous frequency, only rather than fixing the denominator at the sampling rate, the numerator is fixed at 2π.

A graph that plots the mean-interval data points, d_t, against the frame time, t, should be read as, "the expected amount of time it takes to traverse 2π radians of the signal phase at this point in time." A graph of this kind is shown in Figure 6.21A for a recorded tonal. It is difficult to perceive any gradual trend in this series because of noise.

Accumulation and Detrending

The variation of zero crossing interval duration tends to be very small in relation to the mean interval duration. One means of reducing the visual noisiness of the curve in Figure 6.21A is to plot the *cumulative phase traversal*, which we define as the sum of all mean intervals up to and including time step t:

$$\psi_t = \sum_{l=1}^{t} d_l. \tag{6.27}$$

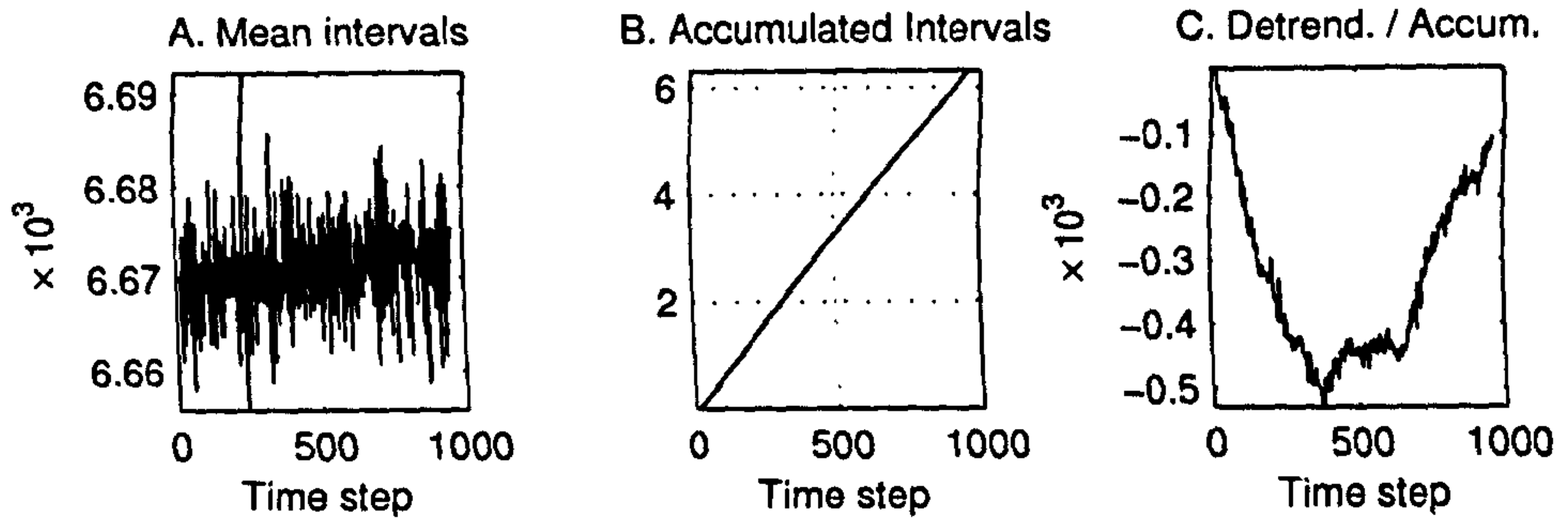


Figure 6.21: A) mean interval data points extracted from a 150 Hz tonal; B) cumulative sum of the data points; C) cumulative sum of the data points with the linear trend removed. A rescaled version of this curve is shown in Figure 6.22D.

The function ψ_t can be interpreted as, "the expected amount of time it is takes to traverse $2\pi t$ radians of the signal phase, based on the evidence of all the frames up to and including time step t."

Figure 6.21B shows the cumulative phase traversal corresponding to the mean interval data points in (A). Frequency variations are now encoded as small changes in the slope of a near-linear trend. The trend grows gradually steeper towards the end, indicating a very slight reduction in component frequency over the course of one minute, although this is almost impossible to see, and extra processing is required to enhance it. The modulation impression around the steady state can be emphasised by subtracting the linear trend away from the cumulative phase traversal[1], resulting in the type of curve shown in Figure 6.21C. (The `detrend` function in MATLAB carries out this operation.)

Comparing Detrended Series

Fluctuations about the steady tonal frequency are perceived more readily in Figure 6.22C than in Figures 6.22A or 6.22B. The detrended cumulative phase traversal curves could form the basis of an operator aid for grouping tonals visually. Figures 6.22B–M plot twelve phase tracks associated with twelve tonal components in a minute-long sonar recording of a merchant vessel. In this case, the tonal components have been picked out manually from the mean power spectrum in Figure 6.22A. (In practice, one would expect a complete system to automate this task using the sort of ZCPA peak tracker described in Section 6.2.) The detrended series have been normalised by their steady frequency.

A casual inspection of Figure 6.22 suggests that the 50 Hz, 150 Hz, 200 Hz and 250 Hz tonals belong together. The 100 Hz and 300 Hz tonals may also belong to this group, as it appears that in both cases large phase jumps have disrupted the accumulation and

[1] The idea of removing the linear trend from a series of phase samples is indebted to QinetiQ (Halse et al., 2005). In that work, the trend is removed from an unwrapped phase track obtained from the DFT (uniform sampling); in this work, the trend is removed from a phase traversal rate track obtained from zero crossing intervals (non-uniform sampling).

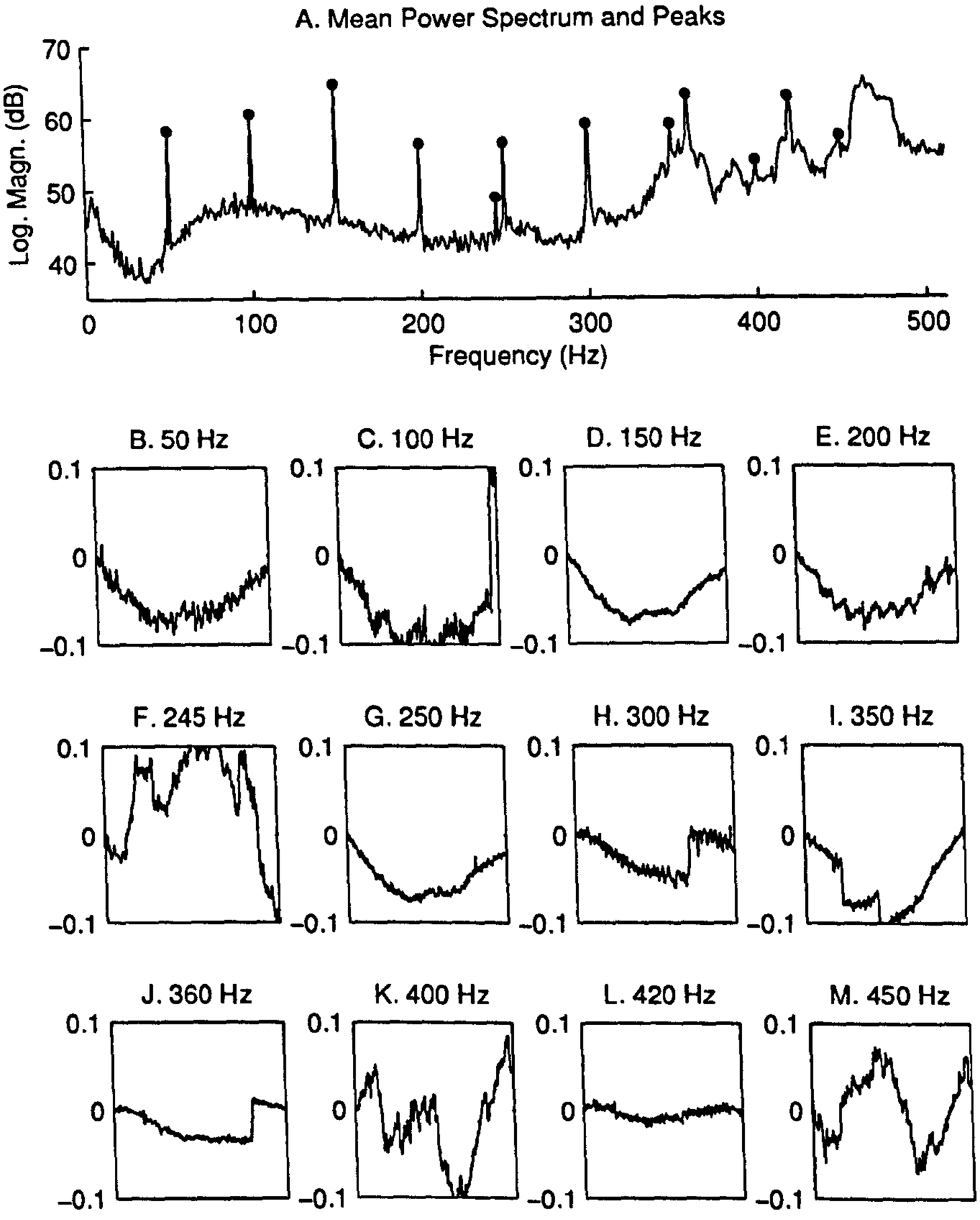


Figure 6.22: A) mean power spectrum with tracked peaks marked using solid circles. B–M) detrended cumulative phase traversal tracks associated with the peaks in (A).

detrending process. Similarly, it appears that the 360 Hz and 420 Hz tonals may have arisen from the same source—perhaps one with a 60 Hz fundamental frequency—although the 360 Hz track is damaged, making an assured judgement difficult. The 300 Hz tonal, which could belong to the 50 Hz or 60 Hz harmonic set, is also interrupted by a phase jump, but its similarity to the 360 Hz track seems to promote the latter association. The 245 Hz, 400 Hz and 450 Hz tracks are too noisy to utilise.

6.5.2 Active Search to Find Similar Tracks

One of the most serious problems facing the grouping of tonals by a passive comparison of phase tracks is that many of these tracks are of very poor quality. Whilst it is possible to identify and remove isolated phase errors in some tracks (*cf.* Figs. 6.22C, H, I and J), other phase tracks are unusable. The most reliable approach to detecting a known, weak signal in white noise is a *matched filter*[1] (Whalen, 1971). In this section, we explore the idea of finding one, clean tonal track and then using robust detection methods to find other tracks like it.

One conceivable problem with the search for a gently-modulated tonal is that a matched filter approach will inevitably find many other tonals present in the recording due to the similarity in their *steady* frequency—albeit with slightly reduced responses. One challenge is to devise a means of distinguishing between signals with the sought-after modulation and signals with different (or no) modulation. Because the null hypothesis encompasses the set of all "other" modulations, and furthermore we lack a model to describe how tonals are expected to behave, we shall choose a pure tonal as the null hypothesis; that is, we shall inquire, "Is this tonal modulated or clean?"

Modulation prominence seems an apposite term for referring to the degree to which a modulated tonal explanation is preferred over a clean tonal explanation. Distinguishing harmonics by FM prominence recalls a study conducted by Marin and McAdams (1991), which demonstrated that human listeners, when presented with an additive mixture of synthetic vowel sounds, assign a higher subjective prominence to a vowel whose harmonics have coherent subaudio frequency modulation. From a mathematical perspective, searching for prominent components in a signal by projecting it onto an overly-rich family of basis functions bears a superficial similarity to the technique of *matching pursuits* (Mallat and Zhang, 1993).

Constructing a Phase Track

Let $\hat{f}_c(t)$ be a relatively clean frequency track estimated using the spline interpolation method of Section 6.3.4. We can construct a *keyed phase track* from $\hat{f}_c(t)$ as follows:

$$\phi_{key}(t) = \frac{2\pi}{\mathrm{avg}(\hat{f}_c)} \int_0^t \hat{f}_c(\tau) d\tau, \tag{6.28}$$

[1] alternatively, a *correlation receiver*.

where avg($\hat{f}_c$) is some measure of the average frequency of the track (e.g., mean or median). The function ϕ_{key} contains the phase track of a unit-frequency signal with the phase variation of f_c impressed upon it. We now wish to search for phase signals at different frequencies, which are modulated in the same manner.

Matched Filter

To determine the extent to which the phase signal $\cos(f\phi_{key})$ appears in the received signal $g(t)$, we suppose that the remainder of the signal is a zero mean additive noise signal, $N(t)$, and then attempt to find the A that minimises

$$E\left\{\int_0^T [A\cos(f\phi_{key}(t)) + N(t) - g(t)]^2\, dt\right\}. \tag{6.29}$$

It can be shown that, provided ϕ_{key} is slow-varying and T is large, the estimate for A which minimises (6.29) has the approximation

$$\hat{A}(f) \approx \frac{2}{T}\int_0^T g(t)\cos(f\phi_{key}(t))\, dt. \tag{6.30}$$

When the initial phase of the signal is unknown, we can estimate the contribution of two components—one in-phase, the other quadrature. This means finding the A_I and A_Q which best explain

$$A_I\cos(f\phi_{key}(t)) + A_Q\sin(f\phi_{key}(t)) + N(t) = g(t),$$

if $g(t)$ is an observed record. In this case, the optimum estimates, in the minimum mean squared error sense, are

$$\hat{A}_I(f) \approx \frac{2}{T}\int_0^T g(t)\cos(f\phi_{key}(t))\, dt \tag{6.31}$$

$$\hat{A}_Q(f) \approx \frac{2}{T}\int_0^T g(t)\sin(f\phi_{key}(t))\, dt. \tag{6.32}$$

When the phase trend is linear, these operations reduce to a frequency-dilated Fourier transform. This assembly forms the basis of the quadrature receiver (Whalen, 1971).

6.5.3 A Non-competitive Explanation

The section above described a method for converting a spline-based frequency track to a nominal phase track and then contructing matched filters to find similarly-modulated phase signals. To test whether this technique could correctly identify whether tonals belonged together, a 70 Hz synthetic tonal complex was mixed with an ocean noise background. Some of the harmonics possessed the unit-frequency-normalised phase track

$$\phi_1(t) = 2\pi t + 2\pi\int_0^t 0.0001\sin(2\pi\cdot 0.02\tau), d\tau; \tag{6.33}$$

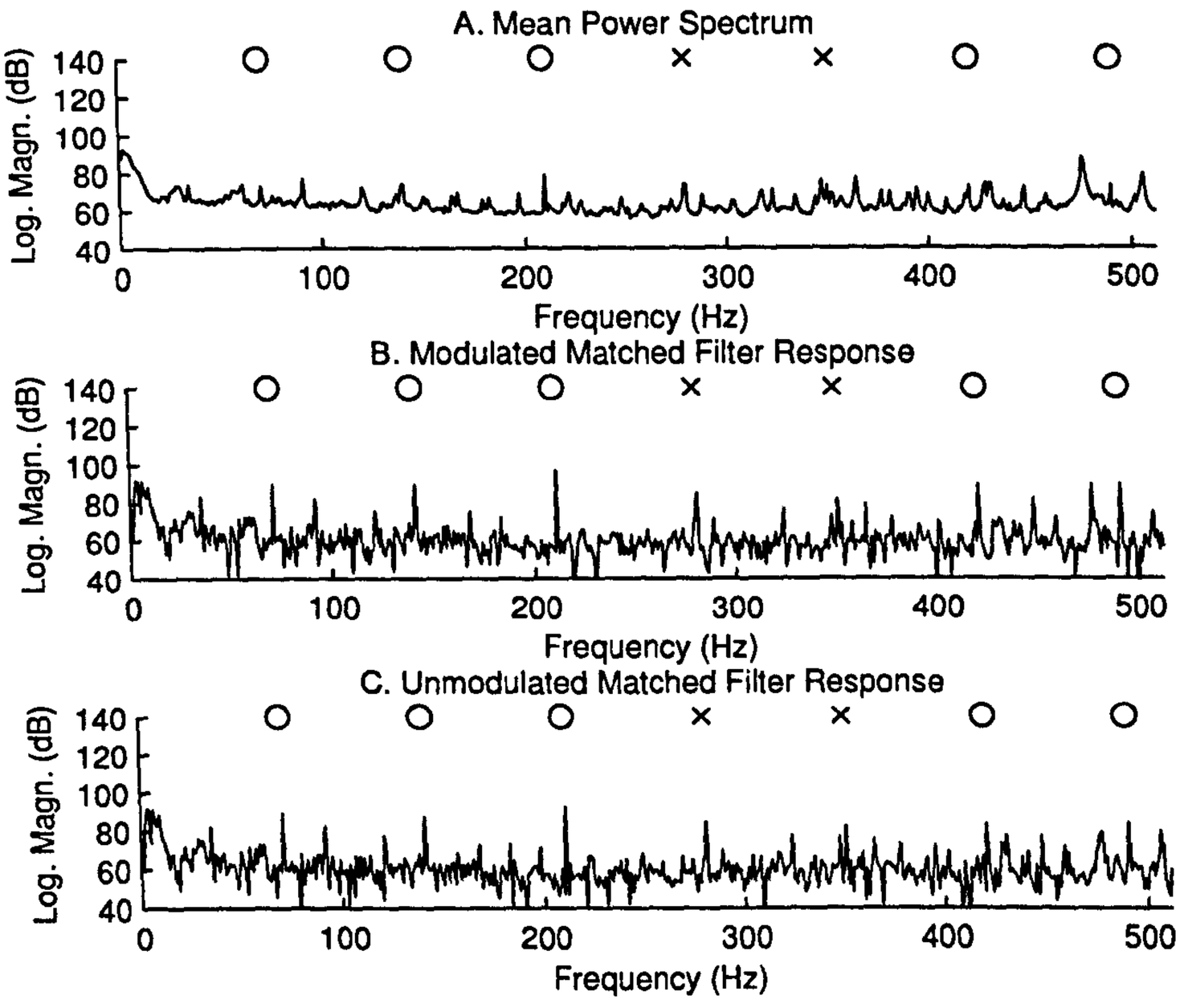


Figure 6.23: Active search for modulated and clean tonals allowing non-competitive explanations. The keyed phase track was extracted from the 210 Hz tonal. A) mean power spectrum measured over one minute; B) response to modulated matched filters; C) response to non-modulated matched filters (Fourier transform).

the others were based on a plain phase track, consistent with a non-modulated tone, i.e., $\phi_2(t) = 2\pi t$. The former group are marked with open circles on Figure 6.23A, the latter group with crosses. In this experiment, the amplitude of the 210 Hz tonal was increased by 3 dB in order to establish a "clean" track to follow. The phase track ϕ_{key} was estimated from this tonal, and one would expect it resemble a noisy version of ϕ_1.

Figure 6.23B plots the function $10 \log_{10} |A_I^2 + A_Q^2|$, when A_I and A_Q are estimated using ϕ_{key}. Figure 6.23C plots the same function, except the estimates A_I and A_Q are generated according to plain phase tracks, making the plot equivalent to a squared magnitude Fourier transform. The intention for this algorithm was that the 1st–3rd, 6th and 7th tracks would appear prominently in (B), and the 4th and 5th tracks would appear prominently in (C), thus effecting a form of separation. However, very little difference can be discerned between the spectra[1].

[1] Some small differences are perceptible when the data are presented on a linear magnitude scale.

The failure of this algorithm can be attributed to the fact that it offers two isolated, *non-competitive explanations* of the signal $g(t)$. Because the phase signals based on the modulated and non-modulated tracks are very similar, the respective matching procedures give near-identical results.

6.5.4 A Competitive Explanation

A *competitive explanation* procedure, as we shall term it, attempts to account for the received signal in terms of an additive mixture containing both types of track, that is,

$$g(t) = N(t) + \sum_{m \in \{key, plain\}} \left[A_{I,m} \cos(f\phi_m(t)) + A_{Q,m} \sin(f\phi_m(t)) \right]. \qquad (6.34)$$

In order to estimate the parameters in this equation, it is helpful to abbreviate the quantities $\cos(f\phi_m(t))$ and $\sin(f\phi_m(t))$ to $\mathrm{c}_m(t)$ and $\mathrm{s}_m(t)$, respectively, and the labels *key* and *plain* to 1 and 2, respectively.

Minimising the mean squared error function,

$$J = E\left\{ \int_0^T \left[A_{I1}\mathrm{c}_1(t) + A_{Q1}\mathrm{s}_1(t) + A_{I2}\mathrm{c}_2(t) + A_{Q2}\mathrm{s}_2(t) + N(t) - g(t) \right]^2 dt \right\},$$

with respect to all A, we arrive at a set of four simultaneous equations,

$$A_{I1}\,\mathrm{c}_1 \cdot \mathrm{c}_1 + A_{Q1}\,\mathrm{s}_1 \cdot \mathrm{c}_1 + A_{I2}\,\mathrm{c}_2 \cdot \mathrm{c}_1 + A_{Q2}\,\mathrm{s}_2 \cdot \mathrm{c}_1 = \mathrm{c}_1 \cdot g \qquad (6.35)$$

$$A_{I1}\,\mathrm{c}_1 \cdot \mathrm{s}_1 + A_{Q1}\,\mathrm{s}_1 \cdot \mathrm{s}_1 + A_{I2}\,\mathrm{c}_2 \cdot \mathrm{s}_1 + A_{Q2}\,\mathrm{s}_2 \cdot \mathrm{s}_1 = \mathrm{s}_1 \cdot g \qquad (6.36)$$

$$A_{I1}\,\mathrm{c}_1 \cdot \mathrm{c}_2 + A_{Q1}\,\mathrm{s}_1 \cdot \mathrm{c}_2 + A_{I2}\,\mathrm{c}_2 \cdot \mathrm{c}_2 + A_{Q2}\,\mathrm{s}_2 \cdot \mathrm{c}_2 = \mathrm{c}_2 \cdot g \qquad (6.37)$$

$$A_{I1}\,\mathrm{c}_1 \cdot \mathrm{s}_2 + A_{Q1}\,\mathrm{s}_1 \cdot \mathrm{s}_2 + A_{I2}\,\mathrm{c}_2 \cdot \mathrm{s}_2 + A_{Q2}\,\mathrm{s}_2 \cdot \mathrm{s}_2 = \mathrm{s}_2 \cdot g, \qquad (6.38)$$

where $a \cdot b$ denotes the inner product, $\int_0^T a(t)b(t)dt$. Once the inner products have been computed, (6.35)–(6.38) can be solved using standard methods to yield the four mixing coefficients $A_{I,key}$, $A_{Q,key}$, $A_{I,pln}$ and $A_{Q,pln}$.

Synthetic Signal Mixed with Recorded Ocean Noise

The active search method has been applied to the mixture of synthetic tones and ocean noise used in the previous section, and the results are set out in Figure 6.24. Now the modulated components—70 Hz, 140 Hz, 210 Hz, 420 Hz and 490 Hz—appear in plot (B) but are missing from plot (C), whilst the unmodulated components—280 Hz, 350 Hz and most of the noise floor—are present in both (B) and (C).

The two spectra are also compared on a linear magnitude scale in Figure 6.24D: the responses of the modulated matched filters are orientated upwards, and the responses of the unmodulated (sinusoidal) matched filters are orientated downwards. Components that are comodulated with the 210 Hz track are absent from the reflection[1]. In addition,

[1] For this reason, the name *Dracula plot* was considered for this kind of display. "This time there could be no error, for the man was close to me, and I could see him over my shoulder. But there was no reflection

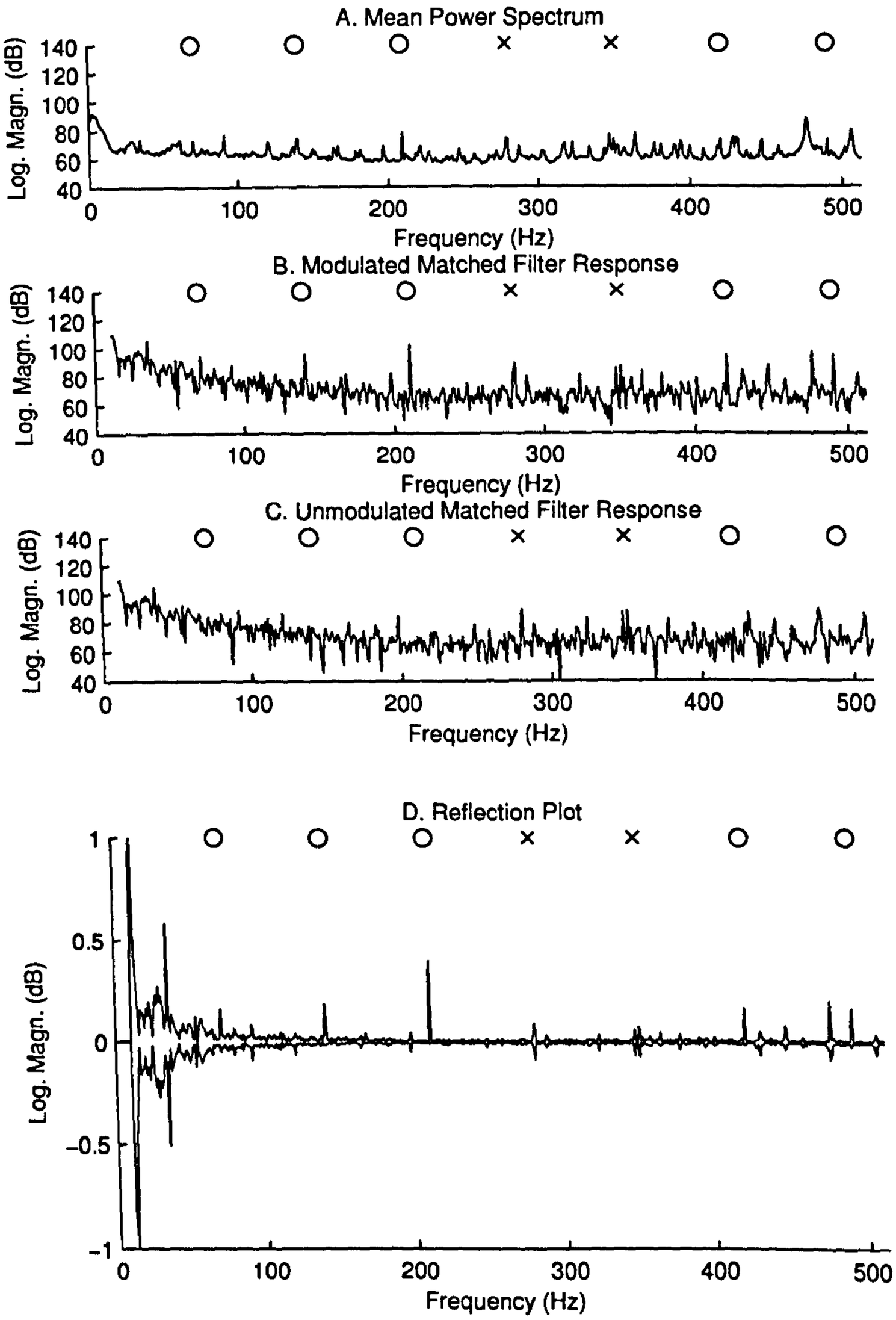


Figure 6.24: Active search for modulated and clean tonals allowing competitive explanations. The keyed phase track was extracted from the 210 Hz tonal. A) mean power spectrum measured over one minute; B) response to modulated matched filters; C) response to non-modulated matched filters (Fourier transform); D) a reflection plot showing the magnitude response to modulated and clean tonals, facing upwards and downwards, respectively.

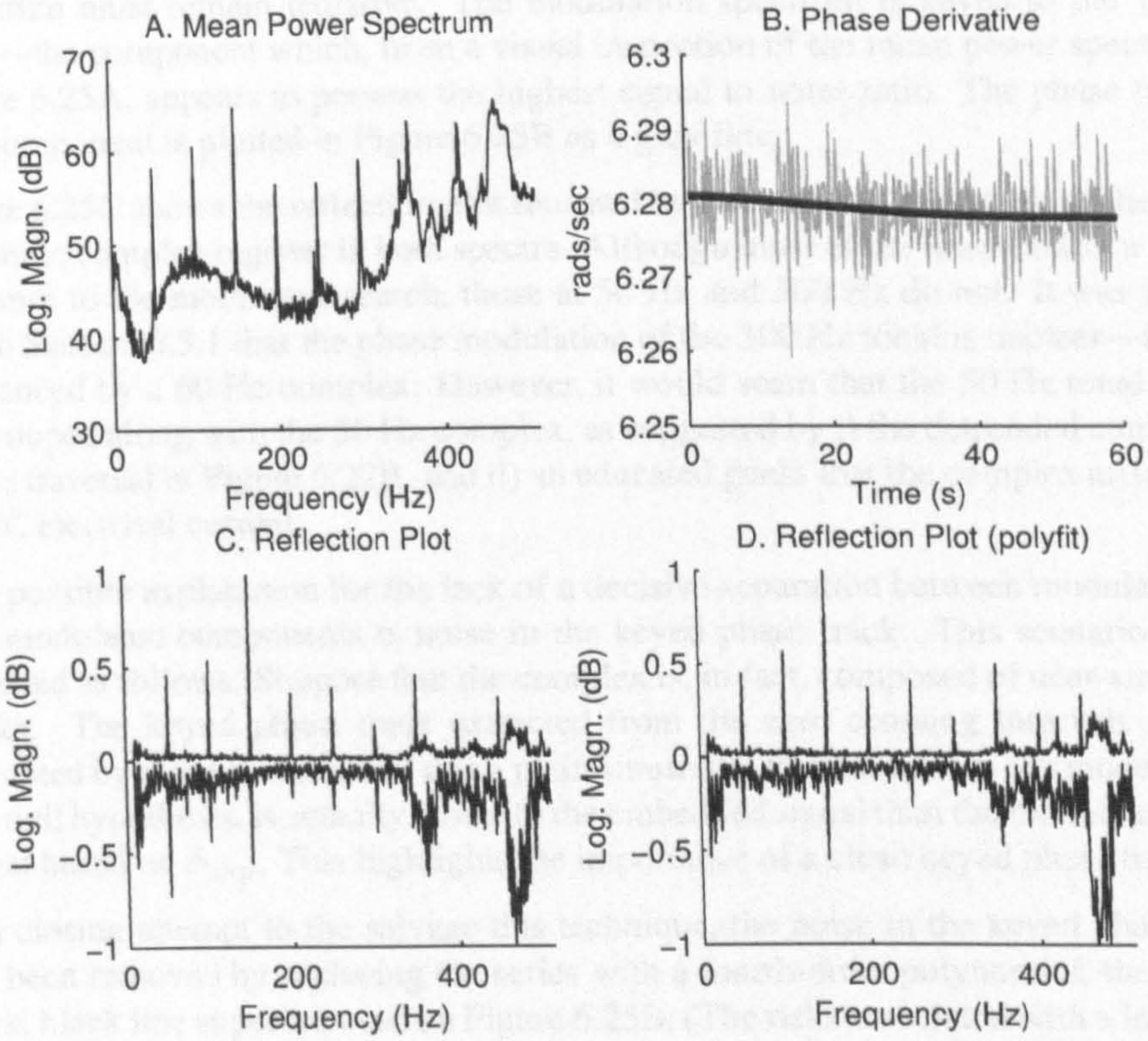


Figure 6.25: Active search for modulated and clean tonals in a minute-long recording of a merchant vessel (with competitive explanations). The keyed phase track was extracted from the 150 Hz tonal. A) mean power spectrum; B) phase track extracted from 150 Hz tonal (grey line) and a fourth-order polynomial fit (heavy, black line); C) a reflection plot obtained for a non-interpolated track; D) a reflection plot obtained for an interpolated track.

the ocean noise recording contains a number of weak tonals, which appear in the reflection. Thus, in principle, even if the constitution of the signal mixture were unknown, this method could be used to determine that the 70 Hz harmonics belonged together.

Recorded Sonar Signal

Lastly, as a final evaluation, we will attempt to use this technique to group the tonals in the merchant vessel recording used in Section 6.5.1. There is no ground truth available for this recording, so judgements concerning the accuracy of the active search

---

of him in the mirror! The whole room behind me was displayed, but there was no sign of a man in it, except myself." (*Dracula*, Chapter II, Bram Stoker)

algorithm must remain tentative. The modulation spectrum is keyed to the 150 Hz tonal—the component which, from a visual inspection of the mean power spectrum in Figure 6.25A, appears to possess the highest signal to noise ratio. The phase track of this component is plotted in Figure 6.25B as a grey line.

Figure 6.25C shows the reflection plot returned by the routine. The tonals of the 50 Hz harmonic complex register in both spectra. Although most of the lines exhibit a greater response to the modulated search, those at 50 Hz and 300 Hz do not. It was pointed out in Section 6.5.1 that the phase modulation of the 300 Hz tonal is unclear—it could influenced by a 60 Hz complex. However, it would seem that the 50 Hz tonal should be grouped along with the 50 Hz complex, as suggested by i) the detrended cumulative phase traversal in Figure 6.22B, and ii) an educated guess that the complex arises from an AC electrical current.

One possible explanation for the lack of a decisive separation between modulated and non-modulated components is noise in the keyed phase track. This scenario can be sketched as follows. Suppose that the complex is, in fact, composed of near-sinusoidal tracks. The keyed phase track extracted from the zero crossing intervals may be corrupted by noise to the extent that a plain sinusoidal track, which in our model serves as a null hypothesis, is actually closer to the embedded signal than the modulated phase signal based on ϕ_{key}. This highlights the importance of a clean keyed phase track.

In a closing attempt to the salvage this technique, the noise in the keyed phase track has been removed by replacing the series with a fourth-order polynomial, shown as a thick, black line superimposed on Figure 6.25B. (The risks associated with a low-order polynomial interpolation are well-documented and were alluded to in Section 6.3.4; however, an inspection of the resulting fit suggests a reasonable approximation in this instance.) The reflection plot generated using the cleaned keyed phase track is provided in Figure 6.25D. Here, the tonals of the 50 Hz harmonic set appear very little in the reflection, if at all, which suggests that some attempt to de-noise the keyed phase track prior to an active search is beneficial.

6.6 Summary

The first goal of this chapter was to assess whether the theoretical work of Chapters 4 and 5 offered any insight into the timing-based processing of the ZCPA-DFT described in Chapter 3. We are now able to assign a random variable to the value that a channel contributes to a ZCPA bin, either with or without peak amplitude weighting, for stationary Gaussian and sinusoidal input signals. By plotting the summary expected contribution of all the filters to all the ZCPA bins, a *mean ZCPA profile* is obtained, which is useful in at least two regards.

The analysis filters tend to generate intervals closer to their centre frequencies, which can give the white ZCPA mean profile an undesirable comb-like appearance, called *ZCPA ripple*. One response to this problem is to widen the bandwidth of the analysis filters; however, doing so reduces post-filter signal-to-noise ratio. The mean ZCPA profile allows a principled trade-off between SNR and ZCPA ripple to be made without the need for experimental white noise testing.

The mean profile can also be used to visualise how a signal or noise background will be manifested in the ZCPA on an absolute scale, which allows a threshold—perhaps one which varies with frequency—to be placed across the ZCPA for signal detection. Unfortunately, in order to perform optimal detection, the variance of the ZCPA bins must be known in addition to the mean. Because this parameter relies on the multiple interval distribution, it is hard to quantify along theoretical lines. This is an important result in itself, however, as it is the work in Chapter 5 which really exposes the difficulty of the problem.

The second goal of the chapter was to investigate the frequency estimation, tracking and grouping of tonal components in sonar recordings using auditory scene analysis principles. The timing-only ZCPA was the representational substrate in which these features were sought. Coarse tracks were formed by joining together closely-spaced peaks in successive ZCPA frames. Two methods for computing the fine frequency estimates "underneath" a coarse track were then discussed. The first approach extended the optimal detection principle from Chapters 4 and 5 to perform optimal estimation. Although this method works, it was deemed impractical, because it requires an unreasonably detailed knowledge of the signal and noise conditions. The second approach fitted a cubic spline through ZCPA data to generate frequency estimates at each sample.

Having chosen the cubic spline method to obtain a set of fine frequency tracks, the problem of grouping tracks together was considered. First, a separate, transient detection pathway was used to reject knots in the spline that were likely to have been corrupted by brief, noisy interruptions. After that, algorithms for grouping tonals were developed, according to two distinct philosophies: the first algorithm involved extracting as many tonals as possible and spotting similar frequency tracks; the remaining algorithms extracted just one or two reliable tonals and actively searched for similar signals.

Chapter

7

Conclusions and Future Work

7.1 Summary

We started by drawing attention to the analogy between a human ear and a passive sonar in functioning as an acoustic receiver and classifier. A closer examination of this analogy revealed that the forms of signal processing effected by these two systems broadly correspond at a number of points, including the mechanisms underlying spatial filtering, ranging, frequency analysis and spectral normalisation, and—more tentatively, in both cases—the measurement of variations in either power or timing to aid signal detection.

The second chapter provided an overview of hearing in humans and placed a particular emphasis on *temporal coding theories* and *auditory scene analysis* (ASA). Temporal coding theories hold that the sequences of neural impulses transmitted to the brain are temporally correlated with the motion of the basilar membrane and are capable of preserving the fine structure of dominant spectral components within a signal. Auditory scene analysis is a conceptual framework for describing and investigating how the auditory system decomposes a sound into sensory elements and then recombines them into perceptual streams.

Several computer models of the auditory periphery and ASA based on temporal coding theories were then surveyed, including the generalised synchrony detector (GSD), ensemble interval histogram (EIH), zero crossings with peak amplitudes (ZCPA), lateral inhibition network (LIN), fine structure spectrogram (FSS), synchrony strands and the autocorrelogram (with other algorithms receiving a passing mention). It was left to the third chapter to decide which of these could best be modified to suit the passive narrowband analysis of sonar signals.

Chapter 3 considered both conventional sonar systems based on power and potential sonar systems based on a temporal analysis. A traditional narrowband spectral analysis displays the signal in the form of a waterfall spectrogram, each row of which is a Fourier magnitude spectrum. An alternative time-frequency representation was sought from one of the timing-based auditory models presented in the preceding chapter, either as a sonar display or as the input format for an ASA model. The ZCPA was chosen on account of its superior resolution, the availability of a robust implementation based on the DFT, and its resemblance to various general-purpose, non-Fourier time-frequency distributions recently proposed[1].

The earliest stages of the project involved a cycle of applying an auditory model to a sonar signal, inspecting its output, observing where a signal was poorly delineated, and finding a way to modify the algorithm to overcome this problem, whilst maintaining its essence. The ZCPA, having survived several iterations of this cycle, appeared to represent simple signals in noise to the naked eye as well as a similarly-configured DFT spectrogram, at which point a new approach was required. Recalling that the theoretical performance of a narrowband sonar display is derived from the DFT, and the magnitude samples of the DFT are, in turn, the output of an envelope detector, it was natural to inquire whether a similar kind of "elementary detector" existed for temporal analyses.

Chapter 4 described the invention of three elementary interval detectors: theoretical models which, when supplied with one zero crossing interval, can discriminate optimally between signal and noise hypotheses, much as a squared-envelope detector can with one sample of the envelope. The hypotheses in this case were wide-sense stationary Gaussian processes. The *sampled interval detector* measures the number of samples between two sign changes in a discrete-time process (i.e., a zero crossing interval) and then uses a Bayesian probability model to decide which hypothesis was most likely to have generated the observation.

It was observed that the performance of the sampled interval detector was overly dependent on a high sample rate, and the *continuous interval detector* was developed as a solution. This detector models the various signal and noise hypotheses in continuous time. The zero crossings of the sampled signal are converted to a continuous time scale using interpolation. The *interpolated interval detector* was proposed as an alternative solution and explicitly models the probability of a random variable consisting of three parts: the (whole) number of samples between two sign changes, and two fractional samples generated by a linear interpolation at each sign change.

The performances of all three elementary interval detectors were compared with that of a squared-envelope detector in a minimum error detection task. (The signal in these experiments was a narrow notch of Gaussian noise.) The interval detectors outperformed the squared-envelope detector when the signal was displaced from the centre of the analysis filter by a certain amount, whereas signals nearer the centre were more reliably detected using power.

[1] e.g., the sparse time-frequency representation, reassigned spectrogram and, of course, the ZCPA itself.

The continuous and interpolated interval detectors performed consistently better than the sampled interval detector on account of the finer sampling of zero crossings. Although these two detectors performed almost identically in every task to which they were applied, the former was much simpler than the latter—both conceptually and in terms of its implementation—so the detectors developed in later chapters were all descendents of the continuous interval detector. The continuous interval detector, despite being the best of the three elementary interval detectors described, remained ill-suited to practical scenarios for at least two reasons.

First, it cannot be configured to optimally detect a pure tone in noise, because the model hypotheses must be wide-sense stationary Gaussian processes. A tone is typically modelled with a constant amplitude and either known phase (in which case it is non-stationary) or uniformly-random phase (in which case it is non-Gaussian). The problem was solved in Chapter 5 by considering the detection of a sinusoid with a fixed but random (Rayleigh-distributed) amplitude and random phase—a Gaussian process. The p.d.f. of the amplitude was then manipulated until the probability density was concentrated around a constant value.

Second, the continuous interval detector can only operate at relatively high SNRs because it processes just one zero crossing interval. One course of action taken to improve the detector was the development of a joint interval-peak detector, which uses the information in both a zero crossing interval and its peak squared amplitude. The joint interval-peak detector matched or outperformed both the elementary interval and squared-envelope detectors under all conditions. Another strategy to improve detection involved the processing of multiple intervals. The successes and failures of multiple interval detection are summarised in the next section.

Chapter 6 returned to the DFT-based ZCPA and investigated whether the theoretical interval detectors could offer any insight into its configuration or further development. Whilst it was possible to discover the mean of a DFT bin, both in the timing-only and amplitude-weighted ZCPAs, a calculation of the variance was intractable, due to the lack of a multiple interval model. This is unfortunate, because knowledge of the mean and variance, combined with post-detection integration of the ZCPA samples, would have led to a Gaussian model. As it is, possessing the mean without the variance still offers some benefits. Specifically, it reveals the average ZCPA noise floor and permits a principled choice of detection threshold. Further, it allows one to view the shape of the mean ZCPA profile in response to noise input and ensure that it is configured to suppress artifacts (e.g., ZCPA ripple).

The latter part of the sixth chapter discussed possible ASA-like processes that could operate above the ZCPA layer. Using the mean profile to choose a suitable narrowband detection threshold for the ZCPA, a peak tracker with a simple continuity constraint was implemented, based on the work of McAulay and Quatieri (1986). The output of this stage was a set of "coarse tracks", which, when superimposed on the ZCPA image, could serve as an operator aid. However, these tracks did not communicate enough fine detail to allow a human user to group them on the basis of low-frequency, shallow modulations.

Given the limited usefulness of coarse tracks for detailed analysis, it was suggested that an operator could manually select a coarse track, and the system would search for a "fine structure track" in the zero crossing information extracted at an earlier stage of the ZCPA. A couple of fine-structure tracking methods were advanced. The first was a model-based solution, which used the interval distributions derived in earlier sections to generate optimal frequency estimates as each zero crossing arrived. A second, data-driven approach formed clumps of zero crossing intervals into the knots of a cubic spline, which was then converted to a frequency track.

The fine frequency tracks obtained using the spline method were incorporated into three algorithms inspired by auditory scene analysis. The first algorithm modelled certain aspects of the "continuity effect", in which a tone is perceived to continue through a brief interruption, such as a noise burst. The algorithm employed a multi-band transient detector, which responds to sudden, energetic, wideband events, to instruct the spline builder to disregard potentially-noisy intervals measured during these periods.

The second and third algorithms were designed to fuse tonals with a common pattern of frequency variation together. The second simply extracted as many fine frequency tracks from the signal as it could find[1] and then plotted the frequency variation of each one so a human inspector could judge which belonged together. Because tonal frequencies vary so gradually, it was necessary to perform some detrending prior to display. The third algorithm, rather than comparing a set of phase tracks, some of which were very noisy, found a single, intense tonal (i.e., one likely to offer a high SNR) and used this as a key to actively search in the signal for similarly-modulated tonals at other frequencies.

[1] In this case, the locations of the tonals were supplied to the algorithm in an *a priori* fashion. However, one would expect an integrated system to highlight features of interest using the coarse tracks in the ZCPA.

7.2 Review of Objectives and Novel Contributions

Each of the objectives from Section 1.3.1 is restated in this section, along with both a cursory reply and a more extended analysis. In some cases, it will be appropriate to point to similar work carried out by others and to highlight the novel aspects of this research.

(1) Can timing-based auditory models be adapted to perform narrowband sonar analysis? What benefits might this offer?

Yes. One benefit is adaptive resolution.

The zero crossings with peak amplitudes (ZCPA) algorithm can be used for narrowband sonar signal analysis. The only significant adaptation required to the original model (of Kim et al., 1999) is a reduction in the bandwidths of the analysis filters—from auditory resolutions (min. 20 Hz–several kHz) to sub-Hertz resolutions. The initial modification of the filterbank is accompanied by changes to the histogram bins, interval window, and so on. It should be noted that these changes are all *quantitative*; the essential behaviour of the ZCPA is unchanged.

Although a bank of model auditory filters (e.g., gammatone) could be constructed with equal centres and bandwidths, it seems more appropriate to use a sliding DFT filterbank. The most obvious benefit of the zero crossing post-processing carried out by the ZCPA is the increase in nominal resolution. The ZCPA resolution can afford to be a few times higher than the filterbank analysis, depending on the overlap in the filter passbands. (See Section 6.1.1 on ZCPA ripple.)

A more subtle benefit of the ZCPA relates to what seems sensible to loosely designate the "time-frequency-SNR uncertainty principle". In the absence of noise, an appropriately-configured ZCPA is able to resolve the time-frequency behaviour of several monocomponent signals to an almost arbitrarily-high degree of precision. The failure to resolve well-separated tonals using extremely fine histogram bins / time steps is due to background noise, which forces us to widen the ZCPA bins to capture variation on a coarser scale. The fact that the ZCPA does not suppress interference amongst closely-spaced components need not be classed as a failure, as the ZCPA will typically represent the whole group as a single, amplitude-modulated component. (We note that the human ear must decide between the "three closely-spaced partials" or "one modulated tone" hypotheses on a regular basis.)

Another helpful feature of time-frequency-SNR uncertainty relates to the precision with which tonals are rendered in the ZCPA image. At a high SNR, the intervals of a steady tonal are relatively undisturbed by the noise, in which case the tonal appears as a sharply-etched, vertical line. At a low SNR, the intervals of a tonal are buffeted by the noise and scattered across a range of histogram bins. In this case, the tonal is manifested in the ZCPA as a less-intense, but broader, swathe. Two advantages emerge from this kind of processing: first, the ZCPA delineates a tonal with a sharpness proportional to the certainty of its frequency; second, and relatedly, a weaker tonal is made easier for the human eye *to detect*, on account of its breadth, whilst a stronger tonal is made easier *to measure*, on account of its narrowness.

(2) How does the performance of an elementary interval detector, which operates on one zero crossing interval, compare to that of a power detector, which operates on one sample of the envelope?

Interval detectors only perform better when the signal is sufficiently displaced from the band centre.

cf. Figs. 4.5, 4.12, 4.18, 4.23, 5.11.

To address this question, both the elementary interval detector and squared-envelope detector were required to detect a narrowband Gaussian process against a white Gaussian noise background. The performance of each detector was measured in terms of the proportion of its decisions that resulted in an error, either a false alarm or false dismissal. In Chapter 4, the parameters of the signal and noise distribution were incorporated into the decision rules, and each detector had only to infer whether the signal was present in the received mixture.

The summary response to this objective is correct; but it is also capable of misleading, as it suggests that the signal might appear anywhere in the band, and the interval detector is more likely to detect it when it is off-centre. In fact, it means that when the squared-envelope and interval detectors are primed to detect a signal at a particular frequency in the analysis band, the interval detector commits fewer errors when the signal is off-centre. Regardless of the signal frequency, the squared-envelope detector always responds to elevated mean energy in the mixture. The interval detector uses zero crossing intervals to refine its response to particular frequency components.

Whence the Interval Distribution?

Objective 2 was written under the assumption that something similar to an elementary interval detector already existed. In fact, references to the notion of optimal detection using zero crossing intervals are surprisingly sparse in the literature, although a great deal of abstract theoretical work is available to anyone wishing to pursue the question. Chief amongst the sources used to produce the elementary interval detector described in this thesis were A. D. Whalen's book, *Detection of Signals in Noise* (1971), B. Kedem's paper, *Spectral analysis and discrimination by zero crossings* (1986), and S. O. Rice's report, *Mathematical Analysis of Random Noise* (1944). Of these, only Rice (1944) suggests a formula to approximate the distribution of the time interval between zero crossings.

Although the circuit taken to arrive at the interval distribution in this work was independent to that of Rice, it is nevertheless indirectly indebted to his work in some places. Aiming to find the distribution governing intervals, the present author utilised Kedem's idea—inherited from Rice—of a sign change in a Gaussian time series to indicate a *zero crossing*, and extended the concept to three signed samples, which, in suitably band-limited conditions, would imply a *zero crossing interval*. To find the probability of an interval event required the derivation of the three-dimensional Gaussian orthant probability, a result which has been known to mathematicians for

over a century[1], and was reworked once again. The limiting process that converted the sampled interval distribution to the continuous interval distribution made an oblique use of Rice's Formula, although this result could itself be found by taking limits.

In answer to the question of whether the continuous interval distribution constitutes a novel contribution, it is certainly fair to say that it is a *novel expression* of the interval distribution, which derives from some of Rice's original ideas, filtered through six decades of later research. Consequently, the interval distribution given in this thesis appears nowhere else in the literature and is markedly different to that of Rice. Other derivations of the interval distribution often mentioned include those of McFadden (1958, 1956) and Longuet-Higgins (1961). The experimental work of Rainal (1962) is also of some relevance.

The studies cited above did not derive or measure the zero crossing interval distribution for the purposes of signal detection. Some later studies explored the possibility of using zero crossings for optimal-like signal detection (e.g., Bae et al., 1996; Higgins, 1980; Bom and Conoly, 1970; Rainal, 1967, 1966), but there is only a passing similarity between these implementations and the elementary interval detectors described in Chapters 4 and 5. A combination of three aspects (italicised)—*optimal* signal detection using the zero crossing *intervals* in *narrowband* signals—appears to constitute a novel approach.

Modulated Gaussian Mixture Models (MGMMs)

A modulated Gaussian mixture model is a parametric description of a one-dimensional function, expressed as a sum of Gaussian-windowed sinusoids (see Section 4.4.4). A suite of `MATLAB` functions was designed to operate on MGMMs, in order to ease the laborious task of manipulating (on paper) the various signals and systems presented in this thesis—especially where a large number of crossterms appear as the result of multiplication.

Technically speaking, no particular one of the routines used to transform MGMMs is novel, and in principle, all of the work presented in this thesis could have been carried out without MGMMs. Nevertheless, the suite as a whole is original and has proven extremely helpful in saving time that would otherwise have been spent writing out (and correcting) Fourier transforms, convolutions, products, squared-magnitudes and so forth.

Although `MATLAB` offers a symbolic tool-box (as does `Maple`), the constrained form of the expressions being handled meant that a set of small, fast, specialised routines operating on one data type was preferred.

[1]Kedem (1980) reproduces this result in isolation but does not connect it to zero crossing intervals.

Placing Sinusoids in the Context of MGMMs

The problem of finding a probability distribution to govern the zero crossing intervals of a sinusoid in filtered Gaussian noise has been tackled before. Cobb (1965), for example, has proposed an approximation in double integral form, which extends the work of Longuet-Higgins and Rice, and relies on series expansions. The approach to modelling this distribution set out in Chapter 5 is, to the knowledge of the author, a novel one, and relies upon a more general result concerning Gaussian processes, which we shall summarise below.

Let $p_X(x \mid \mathcal{G}, \mathcal{A})$ be the probability density function which governs a random quantity X derived from a wide-sense stationary Gaussian process $\mathcal{G}$ by means of the action $\mathcal{A}$. In this case, $\mathcal{A}$ would refer to the extraction of a zero crossing interval. Also, let $\mathcal{G}'$ denote the random process consisting of $\mathcal{G}$ added to a sinusoid with constant amplitude, constant frequency and uniformly-random phase—a non-Gaussian process. Then, it is suggested, there exists a general procedure for approximating the density function $p_X(x \mid \mathcal{G}', \mathcal{A})$.

An explanation of this procedure is given in Section 5.2.5 in the context of zero crossing intervals and MGMMs. This concept is believed to be novel, and researchers in practical signal processing will no doubt find it useful in a wide range of applications.

(3) Is it possible to develop a hybrid detector, which uses both power and timing information? Do a sample of the squared-envelope and a zero crossing interval convey mutually-exclusive or equivalent information?

Yes. Partially-exclusive information.

cf. Figs. 5.13, 5.18.

The naïve joint interval-peak detector simply formed a decision rule based on a product of likelihood ratios. This detector outperformed both the continuous interval detector and squared-envelope detector at all frequencies tried, except those near the centre of the analysis filter, where the squared-envelope detector remained superior. Evidently, the information conveyed in an envelope sample and an interval is not statistically independent, because there are frequencies for which the naïve interval detector has a higher probability of error than the squared envelope detector.

On the other hand, it is also clear that the variables neither convey identical information to each other, nor does the information about one variable consume the other, because there are frequencies for which a naïve joint decision is more accurate than either elementary decision taken alone. We can depict these different possibilities using Venn diagrams, as in Figure 7.1. Of the five interpretations shown in Figure 7.1, only (E) is correct, namely, that the respective measures of information communicated by a sample of the envelope and a zero crossing interval partially overlap.

The *adjusted joint interval-peak detector*, described in Sections 5.3.2–5.3.4 attempts to exploit this overlap to satisify a minimum-error criterion. This likelihood functions in this detector were based on the joint probability density function governing the peaks and intervals of a Gaussian process.

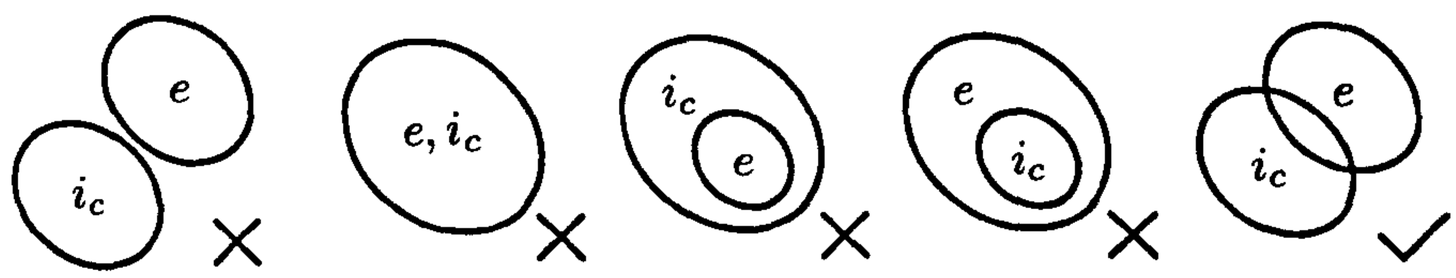


Figure 7.1: Possible information shared by an interval and its peak envelope: A) we cannot know anything about an interval by observing a peak, or *vice versa*; B) if an interval is observed, its peak is known precisely, and *vice versa*; C) an interval tells us everything about its peak but not *vice versa*; D) a peak tells us everything about its interval but not *vice versa*; E) an interval supplies partial information about its peak, and *vice versa*.

A joint distribution of this kind has appeared previously in the literature as it relates to ocean surface waves (Longuet-Higgins, 1983, 1975; Lindgren and Rychlik, 1982). The formula in Chapter 5 was derived by the author independently, as a direct extension of Rice's work and in the context of a particular application (i.e., detection). The contours of the probability density function obtained by the researchers cited above closely resemble Figure 5.17, and these workers also draw attention to the asymmetry exhibited.

(4) Can the elementary interval detector be modified to incorporate multiple interval observations, analogous to a spike train?

In practical terms, only when the intervals are independent.

cf. Figs. 5.29, 5.36 – 5.40.

To understand the principle behind this objective, it is helpful to imagine that one has lost contact with the world outside the head and can only observe the sequence of impulses arriving at the brain at one tonotopic location. What decisions can made on the basis of these observations? Using the present receiver model, the basilar membrane (and perhaps the outer-middle ear) is modelled as a band-pass linear system, and neural transduction is modelled as a zero crossing detector. The elementary interval detector addressed the question of how two successive model spikes relate to Gaussian or sinusoidal input. This objective considers a train of N successive spikes.

As each zero crossing contains two samples, a particular pattern of N intervals is minimally indicated by the signs of $d = 2N+2$ samples. This orthant probability cannot at present be efficiently computed for large dimensionality d without recourse to large series expansions, and a precise solution is known only for $d \leq 3$. An expression for this quantity can be written down from several perspectives, however: as a d-dimensional integral in Cartesian coordinates; as a $(d-1)$-dimensional integral in hyperspherical coordinates; as a d-dimensional solid angle; or, geometrically, in the form of a very-slowly converging sum of simplicial contents. Although numerical solutions for the d-dimensional orthant probability have been published, it was decided that a certain degree of parsimony is desirable in a practical sonar receiver, and alternatives were investigated.

Standard, DFT-based sonar detectors do not process statistically-dependent samples of the envelope, but integrate a series of independent, identically-distributed samples to obtain a Gaussian-distributed test statistic. A similar approach to joint interval-peak detection was taken to satisfy Objective 4. Several independent peak-interval pairs were sampled and averaged, resulting in a two-dimensional Gaussian vector test statistic. To form a decision rule, the means, variances and covariance of the joint interval-peak distribution were computed and duly modified to reflect an average of N independent samples; these were used to parameterise two bivariate Gaussian likelihood functions. The performance of a detector based on this joint statistic exceeded that of a power detector whenever the signal occupied a certain position in the analysis bandwidth (neither near the centre, nor in the filter tail); otherwise the performance matched.

(5) ASA causes a listener to perceive the continuation of a tone which is masked momentarily by noise. Can a similar principle be used in sonar to reconstruct a tonal interrupted by a transient event?

Yes.

cf. Fig. 6.20.

Section 6.4 described how a frequency track can be formed by fitting a cubic spline through the interval data points stored in the ZCPA circular buffers. The mean intervals in each ZCPA frame are used as knots in a spline function, and taking the reciprocal of this function converts interval durations into units of frequency, thus obtaining a frequency track.

To prove the concept of tonal restoration, a synthetic tonal at a constant frequency (401 Hz) was added to a background of recorded ocean noise containing a transient knock. The corresponding frequency track was perturbed during the transient event, as expected, and the repair of the tonal frequency track was subsequently achieved by removing corrupted knots from the spline and recalculating the frequency track. The set of unreliable knots was flagged up by a separate pathway in the algorithm containing a multi-band transient detector.

This stand-alone demonstration can be placed in the more general framework of the interval distribution work that preceded it. When the signal-to-noise ratio in the channel is significantly lowered by the reception of a transient event, the intervals which proceed from the channel are principally determined by the shape of the analysis filter in combination with the colour of the transient spectrum. Under these circumstances, it is best to *disregard* any affected intervals, rather than to assume that they have been corrupted by Gaussian noise (as, e.g., a perturbation analysis would insist).

(6) ASA promotes the fusion of partials exhibiting a common pattern of modulation, especially those in a harmonic relationship. Can a similar principle be used to group engine tonals?

Yes.

cf. Figs. 6.22, 6.23, 6.24, 6.25.

A passive, comparative approach was used to group a set of tonals according to how similarly their rate of phase traversal varied about a steady mean. This so-called *phase traversal rate* is the reciprocal of instantaneous frequency: it relates the time taken to traverse one unit (2π) of phase, rather than the phase traversed in unit time. A similar method for associating tonals, which relied upon coherent variation in the phase of DFT bins, was proposed in a QinetiQ technical report (Halse et al., 2005). One illustration of the principle behind both these algorithms envisages a series of dials, each of which measures the phase of a tonal in relation to its nominal frequency. Any dials that speed up and slow down in unison belong to the same source.

A second means of addressing Objective 6 rested on the idea of "active grouping". Instead of passively comparing the phase trends extracted from potential tonals, an active method uses the phase track of the most reliable tonal to search for other tonals with similar FM against a noise background. Within active methods, a distinction was drawn between non-competitive and competitive explanations.

A non-competitive explanation gives two completely *separate* accounts of the signal. The first is a graph plotting how the total[1] signal energy can be distributed amongst modulated components. The second is a graph showing how the same energy can be divided between non-modulated components (i.e., a Fourier analysis). The intention was that peaks corresponding to comodulated tonals would appear stronger in the first graph than in the second. However, the alternate and null hypotheses are often so similar that this intended difference is unappreciable.

A competitive explanation, to overcome this problem, attempts to reconstruct the signal using an *additive mixture* of modulated and non-modulated components. This technique appeared to be more successful at highlighting the modulated components in a sonar recording than either the non-competitive active or passive comparative approaches, and was applied to both part-synthetic/part-recorded and fully-recorded input signals.

[1] This division of energy is not exact. There will inevitably be some leakage across frequencies, as the modulated basis functions are not quite orthogonal in many cases.

7.3 Future Work

One potential programme of research concerns how the various algorithms described in the thesis could be chained together to form a larger framework for sonar signal analysis. An attractive feature of the proposed system is that it is modular and maintains strict interfaces between modules. One possible modularisation scheme is depicted in Figure 7.2.

Let us consider the zero crossing and peak extraction block as an example. This module relies solely on the output of the filter bank; it does not have access to the raw signal. In addition, its operation is concealed from the layers above. For instance, the circular buffers store the intervals and peaks exiting this block, regardless of the interpolation scheme used to locate crossings, the direction of the crossings detected, the compression function applied to the peaks (and perhaps troughs), and any automatic gain control or lateral inhibition effected.

Similar principles apply to the other blocks. The ZCPA constructs a histogram using the contents of the circular buffers, and the fine structure tracker uses this information to build accurate frequency tracks; but neither block inquires how the circular buffers come to be populated. The coarse peak tracker monitors tonals in the ZCPA without access to the circular buffers, and the human operator monitors the coarse peak tracks, with the option (in this setup) of inspecting the ZCPA and DFT underneath. The only

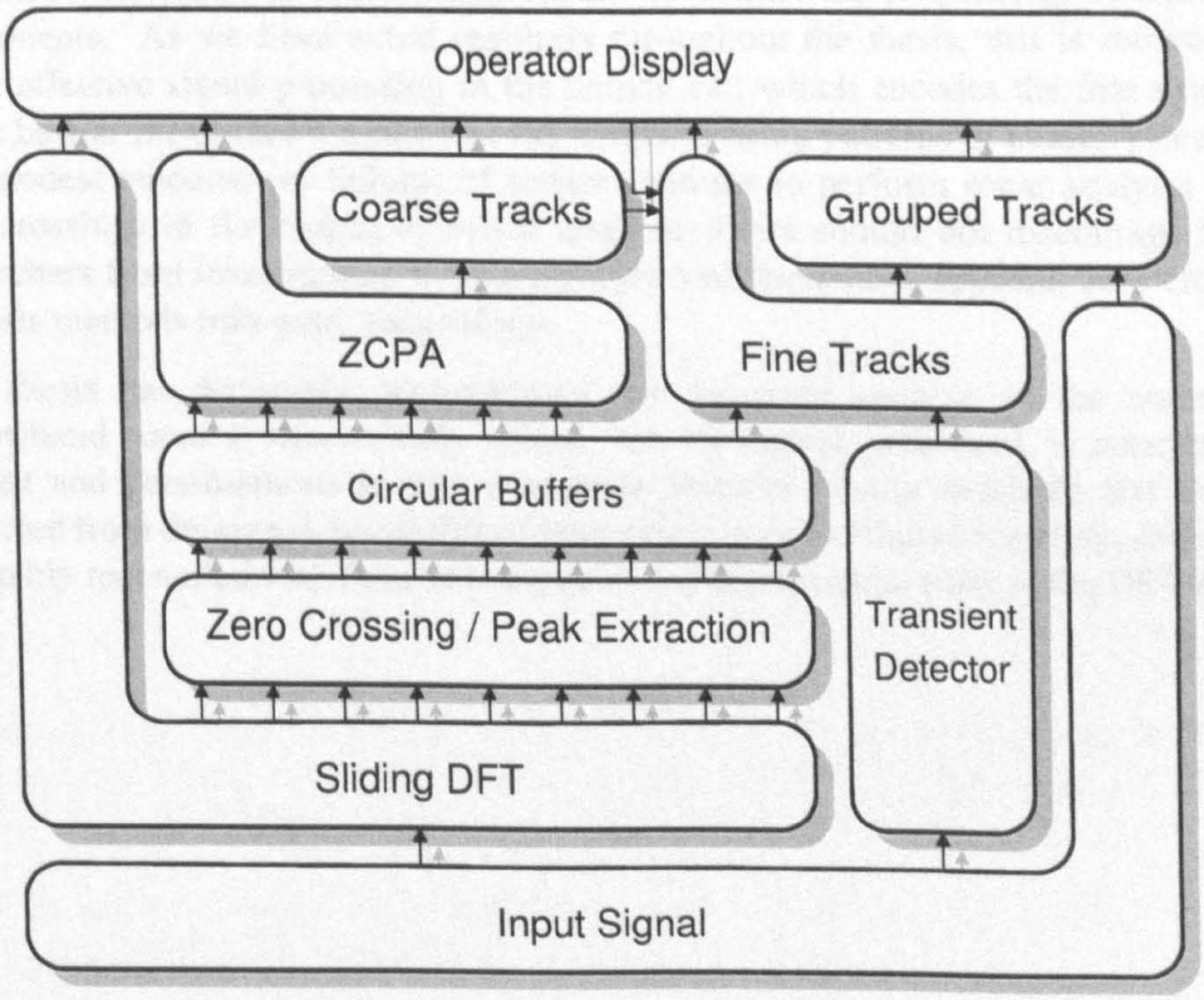


Figure 7.2: An integrated modular system incorporating major aspects of the work carried out in this thesis. Each module only accesses the layer directly beneath.

aspect of this configuration that is not strictly feed-forward is the user's ability to select one or more coarse tracks for surveillance.

Selecting a coarse track has the effect of computing a fine structure track. A cubic spline interpolation algorithm, of the kind proposed in the previous chapter, requires access to the circular buffers to construct knots, as well as a signal from the transient detector to inform which knots, if any, should be discarded. Finally, the grouping algorithm is supplied with a set of fine structure tracks. It selects the cleanest one or few of these and uses them as a pattern to search the input signal for similar components at other frequencies.

7.4 Conclusion

This chapter has provided a summary of the work carried out in this thesis and proposed a scheme for unifying the components described in earlier chapters into an integrated system. The relevance of both low-level mechanisms in the auditory system (such as the basilar membrane and inner hair cells) and high-level organisational principles in hearing (described by auditory scene analysis) have been discussed in relation to sonar. Where possible, computational models of these systems have been adapted to perform sonar analysis and evaluated using synthetic and recorded sonar data.

At present there is a growing interest in non-Fourier time-frequency representations that utilise the phase of narrowband signals to extract the frequencies of dominant components. As we have noted regularly throughout the thesis, this is reminscient of the effective signal processing in the human ear, which encodes the fine structure of the basilar membrane's motion in the temporal firing patterns of auditory neurons. The modest outcome (or failure) of earlier attempts to perform sonar analysis using zero crossings in the output of *broad* analysis filters should not discourage future researchers from investigating the incorporation of these new, subband zero crossing analysis methods into sonar technology.

This thesis has decisively demonstrated that temporal analysis in the context of narrowband sonar is theoretically robust, has biological precedent, is conceptually elegant and parsimonious in that it exploits features readily available and cheaply extracted from the signal, is capable of improving on power-based methods, and can be profitably reconciled with Fourier analysis in implementations such as the DFT-ZCPA.

Bibliography

I. G. Abrahamson. Orthant probabilities for the quadrivariate normal distribution. *Annals of Mathematical Statistics*, 35(4):1685–1703, 1964.

M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, tenth edition, 1972.

W. W. L. Au, A. N. Popper, and R. R. Fay. *Hearing by Whales and Dolphins*. Springer, NY, 2000.

B. M. Ayyub and R. H. McCuen. *Numerical Methods for Engineers*. Prentice Hall, 1995.

J. Bae, Y. Ryu, T. Chang, I. Song, and H.-M. Kim. Nonparametric detection of known and random signals based on zero-crossings. *Signal Processing*, 52:75–82, 1996.

J. S. Bendat and A. G. Piersol. *Random Data: Analysis & Measurement Procedures*. Wiley-Interscience, third edition, 2000.

C. M. Bishop and G. Hinton. *Neural Networks for Pattern Recognition*. Clarendon Press, 1995.

S. Bleeck, P. D. Fox, P. R. White, and N. OMeara. Auditory models and nonlinear filterbanks in underwater auralization. *Journal of the Acoustical Society of America (abstract)*, 123(5):3344, 2008.

N. Bom. Effect of rain on underwater noise level. *Journal of the Acoustical Society of America*, 45(1):150–156, 1969.

N. Bom and B. W. Conoly. Zero-crossing shift as a detection method. *Journal of the Acoustical Society of America*, 47(5):1408–1411, 1970.

A. S. Bregman. *Auditory Scene Analysis: the perceptual organisation of sound*. MIT Press, 1990.

A. S. Bregman and G. Dannenbring. The effect of continuity on auditory stream segregation. *Perception & Psychophysics*, 13:308–312, 1973.

G. J. Brown. *Computational Auditory Scene Analysis: A Representational Approach*. PhD thesis, University of Sheffield, 1992.

G. J. Brown and M. P. Cooke. Computational auditory scene analysis. *Computer Speech and Language*, 8:297–336, 1994.

J. F. Brugge, D. J. Anderson, J. E. Hind, and J. E. Rose. Time structure of discharges in single auditory nerve fibers of the squirrel monkey in response to complex periodic sounds. *Journal of Neurophysiology*, 32:386–401, 1969.

J. F. Brugge and M. M. Merzenich. Responses of neurons in auditory cortex of the macaque monkey to monaural and binaural stimulation. *Journal of Neurophysiology*, 36:1138–1158, 1973.

W. S. Burdic. *Underwater Acoustic Systems Analysis*. Englewood Cliffs, NJ; Prentice-Hall, first edition, 1984.

V. Ciocca and A. S. Bregman. Perceived continuity of gliding and steady-state tones through interrupting noise. *Perception & Psychophysics*, 42:476–484, 1987.

L. Cohen. *Time-frequency Analysis*. Prentice Hall PTR, 1995.

M. P. Cooke. *Modelling auditory processing and organisation*. PhD thesis, University of Sheffield, 1991/1993.

M. P. Cooke and D. P. W. Ellis. The auditory organization of speech and other sources in listeners and computational models. *Speech Communication*, 35:141–177, 2001.

J. W. Cooley and J. W. Tukey. An algorithm for the machine computation of complex fourier series. *Mathematics of Computation*, 19:297–301, 1965.

H. R. Dajani, W. Wong, and H. Kunov. Fine structure spectrography and its application in speech. *Journal of the Acoustical Society of America*, 117(6): 3902–3918, 2005.

T. Dau, D. Püschel, and A. Kohlrausch. A quantitative model of the "effective" signal processing in the auditory system. I. model structure. *Journal of the Acoustical Society of America*, 99(6):3615–3622, 1996.

F. N. David. A note on the evaluation of the multivariate normal integral. *Biometrika*, 40:458–459, 1953.

R. L. Dawe. Detection threshold modelling explained. Technical Report DSTO-TR-0586, DSTO Aeronautical and Maritime Research Laboratory, 1997.

E. de Boer. Travelling waves and cochlear resonance. *Scandinavian Audiology Supplement*, 9:17–33, 1979.

E. de Boer and H. R. de Jongh. On cochlear encoding: Potentialities and limitations of the reverse-correlation technique. *Journal of the Acoustical Society of America*, 63(1):115–135, 1978.

A. de Cheveigné. Pitch perception models. In C.J. Plack, A.J. Oxenham, R.R. Fay, and A.N. Popper, editors, *Pitch*. Springer, 2005.

L. Deng and H. Sheikhzadeh. Use of temporal codes computed from a cochlear model for speech recognition. In S. Greenberg and W.A. Ainsworth, editors, *Listening to Speech: An Auditory Perspective*. Lawrence Erlbaum Associates, 2006.

P. Depalle, G. Garcia, and X. Rodet. Tracking of partials for additive sound synthesis using hidden markov models. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 225–228, 1993.

J. J. Eggermont. Temporal modulation transfer functions for AM and FM stimuli in cat auditory cortex. effects of carrier type, modulating waveform and intensity. *Hearing Research*, 74:51–66, 1994.

D. P. W. Ellis. *Prediction-driven computational auditory scene analysis*. PhD thesis, MIT, 1996.

S. Finger. *Origins of Neuroscience: A History of Explorations Into Brain Function*. Oxford University Press US, 2001.

S. A. Fulop and K. Fitz. Algorithms for computing the time-corrected instantaneous frequency (reassigned) spectrogram, with applications. *Journal of the Acoustical Society of America*, 119(1):360–371, 2006.

T. J. Gardner and M. O. Magnasco. Sparse time-frequency representations. *Proceedings of the National Academy of Sciences*, 103(16):6094–6099, 2006.

W. W. Gaver. What in the world do we hear? an ecological approach to auditory perception. *Ecological Psychology*, 5(1):1–29, 1993.

O. Ghitza. Temporal non-place information in the auditory-nerve firing patterns as a front-end for speech recognition in a noisy environment. *Journal of Phonetics*, 16: 109–123, 1988.

O. Ghitza. Auditory models and human performance in tasks related to speech coding and speech recognition. *IEEE Transactions on Speech and Audio Processing*, 2(1): 115–131, 1994.

A. C. Gilbert, M. J. Strauss, and J. A. Tropp. A tutorial on fast fourier sampling. *IEEE Signal Processing Magazine*, 25(2):57–66, 2008.

B. R. Glasberg and B. C. J. Moore. Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47:103–138, 1990.

B. R. Glasberg and B. C. J. Moore. Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise. *Journal of the Acoustical Society of America*, 108(5):2318–2328, 2000.

G. Goertzel. An algorithm for the evaluation of finite trigonometric series. *American Mathematical Monthly*, 65:34–35, 1958.

B. Gold and N. Morgan. *Speech and Audio Signal Processing: Processing and Perception of Speech and Music*. John Wiley & Sons, Inc., 2000.

R. L. Goode, M. Killion, K. Nakamura, and S. Nishihara. New knowledge about the function of the human middle ear: development of an improved analog model. *American Journal of Otolaryngology*, 15(2):145–154, 1994.

S. Greenberg and W. A. Ainsworth. Speech processing in the auditory system: An overview. In S. Greenberg, W. A. Ainsworth, A.N. Popper, and R.R. Fay, editors, *Speech Processing in the Auditory System*. Springer, first edition, 2006.

J. M. Grey. Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, 61(5):1270–1277, 1977.

A. Grigorakis. Application of detection theory to the measurement of the minimum discernable signal for a sinusoid in gaussian noise displayed on a lofargram. Technical Report DSTO-TR-0568, DSTO Aeronautical and Maritime Research Laboratory, 1997.

M. Hajja and P. Walker. The measure of solid angles in n-dimensional euclidean space. *International Journal of Mathematical Education in Science and Technology*, 33(5):725–800, 2002.

J. W. Hall, M. P. Haggard, and M. A. Fernandes. Detection in noise by spectro-temporal pattern analysis. *Journal of the Acoustical Society of America*, 76 (1):50–56, 1984.

P. Halse, A. McLean, P. Tindell, and R. Wilcox. Narrow-band phase analysis and tonal association in support of data fusion (draft). Technical report, QinetiQ, July 2005.

F. J. Harris. On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE*, 66(1):51–83, 1978.

R. C. Higgins. The utilization of zero-crossing statistics for signal detection. *Journal of the Acoustical Society of America*, 67(5):1818–1820, 1980.

T. Irino and R. D. Patterson. A time-domain, level-dependent auditory filter: The gammachirp. *Journal of the Acoustical Society of America*, 101(1):412–419, 1997.

ISO. Normal equal-loudness level contours for pure tones under free-field listening conditions. Technical Report ISO-226:2003, International Organization for Standardization, Geneva, Switzerland, 2003.

E. Jacobsen and R. Lyons. The sliding DFT. *IEEE Signal Processing Magazine*, 20 (2):74–80, 2003.

E. Javel. Coding of AM tones in the chinchilla auditory nerve: Implications for the pitch of complex tones. *Journal of the Acoustical Society of America*, 68:133–146, 1980.

P. X. Joris, L. H. Carney, P. H. Smith, and T. C. Yin. Enhancement of neural synchronization in the anteroventral cochlear nucleus. i. responses to tones at the characteristic frequency. *Journal of Neurophysiology*, 71(3):1022–36, 1994.

J. M. Kates. A time-domain digital cochlear model. *IEEE Transactions on Signal Processing*, 39(12):2573–2592, 1991.

S. M. Kay and R. Sudhaker. A zero crossing-based spectrum analyzer. *IEEE Transactions on Acoustics Speech and Signal Processing*, 34(1):96–104, 1986.

B. Kedem. *Binary time series*. Marcel Dekker, Inc., 270 Madison Avenue, New York, New York, 10016, 1980.

B. Kedem. Spectral analysis and discrimination by zero-crossings. *Proceedings of the IEEE*, 74:1477–1493, 1986.

S. M. Khanna and M. C. Teich. Spectral characteristics of the responses of primary auditory-nerve fibers to amplitude-modulated signals. *Hearing Research*, 39(1-2): 143–157, 1989.

N. Y. S. Kiang. Processing of speech by the auditory nervous system. *Journal of the Acoustical Society of America*, 68(3):830–835, 1980.

N. Y. S. Kiang and E. C. Moxon. Tails of tuning curves of auditory-nerve fibers. *Journal of the Acoustical Society of America*, 55(3):620–630, 1974.

D.-S. Kim, S.-Y. Lee, and R.-M. Kil. Auditory processing of speech signals for robust speech recognition in real-world noisy environments. *IEEE Transactions on Speech and Audio Processing*, 7(1):55–69, 1999.

L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders. *Fundamentals of Acoustics*. John Wiley and Sons, Inc., fourth edition, 2000.

I. P. Kirsteins, S. K. Mehta, and J. Fay. Separation and fusion of overlapping underwater sound streams. In *Proceedings of EUSIPCO 2000, Tampere*, volume 2, pages 1109–1113, 2000.

V. O. Knudsen, R. S. Alford, and J. W. Emling. Underwater ambient noise. *Journal of Marine Research*, 7:410–429, 1948.

K. Kodera, R. Gendrin, and C. De Villedary. Analysis of time-varying signals with small BT values. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 26(1):64–76, 1978.

K. Kodera, C. De Villedary, and R. Gendrin. A new method for the numerical analysis of non-stationary signals. *Physics of The Earth and Planetary Interiors*, 12(2–3): 142–150, 1976.

K. Koffka. *Principles of Gestalt Psychology*. Harcourt Brace, New York, 1935.

R. Kumaresan and A. Rao. Model-based approach to envelope and positive instantaneous frequency estimation of signals with speech applications. *Journal of the Acoustical Society of America*, 105(3):1912–1924, 1999.

R. Kumaresan and Y. Wang. On representing signals using only timing information. *Journal of the Acoustical Society of America*, 110(5):2421–2439, 2001.

M. Lagrange, S. Marchand, and J. Rault. Using linear prediction to enhance the tracking of partials. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 241–244, 2004.

M. Lasky. Review of undersea acoustics to 1950. *Journal of the Acoustical Society of America*, 61(2):283–297, 1977.

M. C. Liberman. Auditory-nerve response from cats raised in a low-noise chamber. *Journal of the Acoustical Society of America*, 63(2):442–455, 1978.

J. C. R. Licklider. A duplex theory of pitch perception. *Experimentia*, 7:128–133, 1951.

G. Lindgren and I. Rychlik. Wave characteristic distributions for gaussian waves—wave-length, amplitude and steepness. *Ocean Engineering*, 9(5):411–432, 1982.

B. F. Logan. Information in the zero crossings of band-pass signals. *Bell Systems Technical Journal*, 56:487–510, 1977.

M. S. Longuet-Higgins. The distribution of intervals between zeros of a stationary random function. *Philosophical Transactions of the Royal Society of London, Series A*, 254:557–559, 1961.

M. S. Longuet-Higgins. On the joint distribution of the periods and amplitudes of sea waves. *Journal of Geophysical Research*, 80:2688–2694, 1975.

M. S. Longuet-Higgins. On the joint distribution of wave periods and amplitudes in a random wave field. *Philosophical Transactions of the Royal Society of London, Series A*, 389:241–258, 1983.

E. A. Lopez-Poveda and R. Meddis. A human nonlinear cochlear filterbank. *Journal of the Acoustical Society of America*, 110(6):3107–3118, 2001.

T. J. Lynch, III, V. Nedzelnitsky, and W. T. Peake. Input impedance of the cochlea in cat. *Journal of the Acoustical Society of America*, 72(1):108–130, 1982.

R. F. Lyon. A computational model of filtering, detection and compression in the cochlea. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 7, pages 1282–1285, 1982.

S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, 1993.

C. M. H. Marin and S. McAdams. Segregation of concurrent sounds. II: Effects of spectral envelope tracing, frequency modulation coherence, and frequency modulation width. *Journal of the Acoustical Society of America*, 89(1):341–351, 1991.

G. Marsaglia. Choosing a point from the surface of a sphere. *Annals of Mathematical Statistics*, 43:645–646, 1972.

R. J. McAulay and T. F. Quatieri. Speech analysis/synthesis based on a sinusoidal representation. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 34 (4):744–754, 1986.

J. McFadden. The axis-crossing intervals of random functions. *IRE Transactions on Information Theory*, 2(4):146–150, 1956.

J. McFadden. The axis-crossing intervals of random functions–II. *IRE Transactions on Information Theory*, 4(1):14–24, 1958.

R. Meddis. Simulation of mechanical to neural transduction in the auditory receptor. *Journal of the Acoustical Society of America*, 79:702–711, 1986.

R. Meddis. Simulation of auditory-neural transduction: Further studies. *Journal of the Acoustical Society of America*, 83(3):1056–1063, 1988.

R. Meddis and M. J. Hewitt. Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. *Journal of the Acoustical Society of America*, 89(6):2866–2882, 1991.

D. Mellinger. *Event formation and separation in musical sound.* PhD thesis, Stanford University, 1991.

D. H. Mershon and J. N. Bowers. Absolute and relative cues for the auditory perception of egocentric distance. *Perception*, 8:311–322, 1979.

G. Meyer. Anatomical and physiological bases of speech perception. In S. Greenberg and W. A. Ainsworth, editors, *Listening to Speech: An Auditory Perspective*. Lawrence Erlbaum Associates, 2006.

G. A. Miller and J. C. R. Licklider. The intelligibility of interrupted speech. *Journal of the Acoustical Society of America*, 22(2):167–173, 1950.

A. R. Møller. *Hearing: Anatomy, Physiology and Disorders of the Auditory System.* Academic Press, second edition, 2006.

B. C. J. Moore. *An Introduction to the Psychology of Hearing*. Academic Press, fifth edition, 2004.

M. E. Muller. A note on a method for generating points uniformly on n-dimensional spheres. *Communications of the ACM*, 2(4):19–20, 1959.

T. Nakatani. *Computational Auditory Scene Analysis based on Residue-Driven Architecture and its Applications to Mixed Speech Recognition*. PhD thesis, Kyoto University, 2002.

K. E. Nilsen and I. J. Russell. Spatial and temporal representation of a tone on the guinea pig basilar membrane. *Proceedings of the National Academy of Sciences*, 97 (22):11751–11758, 2000.

A. V. Oppenheim and R. W. Schafer. *Discrete-Time Signal Processing*. Prentice-Hall, 1989.

P. B. Ostergaard. Implementation details of a computation model of the inner hair-cell/auditory-nerve synapse. *Journal of the Acoustical Society of America*, 87 (4):1813–1816, 1990.

A. R. Palmer and I. J. Russell. Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells. *Hearing Research*, 24 (1):1–15, 1986.

H.-M. Park and R. M. Stern. Spatial separation of speech signals using continuously-variable masks estimated from comparisons of zero crossings. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 4:1165–1168, 2006.

T. W. Parks and B. A. Weisburn. Classification of whale and ice sounds with a cochlear model. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages 481–484, 1992.

T. W. Parsons. Separation of speech from interfering speech by means of harmonic selection. *Journal of the Acoustical Society of America*, 60(4):911–918, 1976.

R. Patterson, I. Nimmo-Smith, J. Holdsworth, and P. Rice. Spiral VOS final report part A: The auditory filter bank. Internal Report 2341, MRC Applied Psychology Unit, Cambridge, UK, 1988.

R. D. Patterson. Auditory filter shapes derived with noise stimuli. *Journal of the Acoustical Society of America*, 59(3):640–654, 1976.

R. D. Patterson, M. H. Allerhand, and C. Giguère. Time-domain modelling of peripheral auditory processing: A modular architecture and software platform. *Journal of the Acoustical Society of America*, 98(4):1890–1894, 1995.

P. Z. Peebles, Jr. *Probability, Random Variables and Random Signal Principles*. McGraw-Hill Inc., third edition, 1993.

D. P. Phillips and S. E. Hall. Responses of single neurons in cat auditory cortex to time-varying stimuli: linear amplitude modulations. *Experimental Brain Research*, 67(3):479–492, 1987.

J. O. Pickles. *An Introduction to the Physiology of Hearing*. Academic Press, second edition, 1988.

D. Pralong and S. Carlile. The role of individualized headphone calibration for the generation of high fidelity virtual auditory space. *Journal of the Acoustical Society of America*, 100(6):3785–3793, 1996.

W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 1992.

C. E. Pykett and D. J. Holland Smith. Narrow-band phase analysis and tonal association in support of data fusion (draft). Technical report, QinetiQ, March 2000.

L. R. Rabiner. A tutorial on the hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

A. J. Rainal. Zero-crossing intervals of gaussian processes. *IRE Transactions on Information Theory*, 8(6):372–378, 1962.

A. J. Rainal. Zero-crossing principle for detecting narrow-band signals. *IEEE Transactions on Instrumentation and Measurement*, IM-15(1–2):38–43, 1966.

A. J. Rainal. Another zero-crossing principle for detecting narrow-band signals. *IEEE Transactions on Instrumentation and Measurement*, IM-16(2):134–138, 1967.

J. M. Ribando. Measuring solid angles beyond dimension three. *Discrete & Computational Geometry*, 36(3):479–487, 2006.

S. O. Rice. Mathematical analysis of random noise. *Bell Systems Technical Journal*, 23:282–332, 1944.

L. Robles and M. A. Ruggero. Mechanics of the mammalian cochlea. *Physiological Reviews*, 81:1305–1352, 2001.

I. Rychlik. Joint distribution of successive zero crossing distances for stationary gaussian processes. *Journal of Applied Probability*, 24(2):378–385, 1987.

M. B. Sachs, C. C. Blackburn, and E. D. Young. Rate-place and temporal-place representations of vowels in the auditory nerve and anteroventral cochlear nucleus. *Journal of Phonetics*, 16:37–53, 1988.

M. B. Sachs and N. Y. S. Kiang. Two-tone inhibition in auditory-nerve fibers. *Journal of the Acoustical Society of America*, 43(5):1120–1128, 1968.

M. B. Sachs, B. J. May, G. S. Le Prell, and R. D. Hienz. Adequacy of auditory-nerve rate representations of vowels: Comparison with behavioural measures in cat. In S. Greenberg and W. A. Ainsworth, editors, *Listening to Speech: An Auditory Perspective*. Lawrence Erlbaum Associates, 2006.

M. B. Sachs and E. D. Young. Encoding of steady-state vowels in the auditory nerve: Representation in terms of discharge rate. *Journal of the Acoustical Society of America*, 66(2):470–479, 1979.

D. Schofield. Visualations of speech based on a model of the peripheral auditory system. Technical Report NPL Report DITC 62/85, National Physical Laboratory, 1985.

G. P. Schoonveldt and B. C. J. Moore. Comodulation masking release (CMR) as a function of masker bandwidth, modulator bandwidth and signal duration. *Journal of the Acoustical Society of America*, 82(1):273–281, 1989.

M. R. Schroeder. Period histogram and product spectrum: New methods for fundamental-frequency measurement. *Journal of the Acoustical Society of America*, 43:829–834, 1968.

M. R. Schroeder. An integrable model for the basilar membrane. *Journal of the Acoustical Society of America*, 53(2):429–434, 1973.

J. A. Scrimger, D. J. Evans, G. A. McBean, D. M. Farmer, and B. R. Kerman. Underwater noise due to rain, hail, and snow. *Journal of the Acoustical Society of America*, 81(1):79–86, 1987.

S. Chandra Sekhar and T. V. Sreenivas. Auditory motivated level-crossing approach to instantaneous frequency estimation. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 53:1450–1462, 2005.

R. H. Self. Asymptotic expansion of integrals. In M.C.M. Wright, editor, *Lecture Notes on the Mathematics of Acoustics*. Imperial College Press, 2005.

S. Seneff. A joint synchrony/mean-rate model of auditory speech processing. *Journal of Phonetics*, 16:55–76, 1988.

S. A. Shamma. Speech processing in the auditory system I: The representation of speech sounds in the responses of the auditory nerve. *Journal of the Acoustical Society of America*, 78(5):1612–1621, 1985a.

S. A. Shamma. Speech processing in the auditory system II: Lateral inhibition and the central processing of speech evoked activity in the auditory nerve. *Journal of the Acoustical Society of America*, 78(5):1622–1632, 1985b.

K. S. Shanmugan and A. M. Breipohl. *Random Signals: Detection, Estimation and Data Analysis*. John Wiley and Sons, Inc., 605 Third Avenue, New York, New York 10158, 1988.

M. Slaney. Lyon's cochlear model. Technical Report Apple Tech. Report #13, Apple Inc., 1988.

M. Slaney. Auditory toolbox: A matlab toolbox for auditory modeling work. Technical Report Apple Tech. Report #45, Apple Inc., 1994.

M. Slaney and R. F. Lyon. A perceptual pitch detector. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 357–360, 1990.

T. V. Sreenivas and R. J. Niederjohn. Zero-crossing based spectral analysis and svd spectral analysis for formant frequency estimation in noise. *IEEE Transactions on Signal Processing*, 40(2):282–293, 1992.

T. A. Stroffregen and J. B. Pittenger. Human echolocation as a basic form of perception and action. *Journal of Ecological Psychology*, 7(3):181–216, 1995.

Q. Summerfield and P. F. Assmann. Perception of concurrent vowels: Effects of harmonic misalignment and pitch-period asynchrony. *Journal of the Acoustical Society of America*, 89(3):1364–1377, 1991.

C. J. Sumner, E. A. Lopez-Poveda, L. P. O'Mard, and R. Meddis. A revised model of the inner-hair cell and auditory-nerve complex. *Journal of the Acoustical Society of America*, 111(5):2178–2188, 2002.

I. Tasaki. Nerve impulses in individual auditory nerve fibers of guinea pig. *Journal of Neurophysiology*, 17(2):97–122, 1954.

A. Teolis and S. Shamma. Classification of transient signals via auditory representations. Technical Report TR 91-99, University of Maryland, Systems Research Center, 1991.

N. P. McA. Todd. The auditory "primal sketch": A multiscale model of rhythmic grouping. *Journal of New Music Research*, 23:25–70, 1994.

S. Tucker. *An ecological approach to the classification of transient underwater acoustic events: Perceptual experiments and auditory models*. PhD thesis, University of Sheffield, 2003.

S. Tucker and G. J. Brown. Classification of transient sonar sounds using perceptually motivated features. *IEEE Journal of Oceanic Engineering*, 30(3), 2005.

M. Unoki and M. Akagi. A method of signal extraction from noisy signal based on auditory scene analysis. *Speech Communication*, 27(3–4):261–279, 1999.

R. J. Urick. Signal excess and detection probability of fluctuating sonar signals in noise. *Journal of the Acoustical Society of America*, 60:S18, 1976.

L. P. A. S. van Noorden. *Temporal Coherence in the Perception of Tone Sequences*. PhD thesis, Eindhoven University of Technology, 1975.

H. B. Voelcker. Towards a unified theory of modulation part I: Phase-envelope relationships. *Proceedings of the IEEE*, 54:340–354, 1966.

G. von Békésy. The variation of phase along the basilar membrame with sinusoidal vibrations. *Journal of the Acoustical Society of America*, 19(3):452–460, 1947.

G. von Békésy and W. A. Rosenblith. The early history of hearing—observations and theories. *Journal of the Acoustical Society of America*, 20(6):727–748, 1948.

R. A. Wagstaff. The Wagstaffs integration silencing processor filter: A method for exploiting fluctuations to achieve improved sonar signal processor performance. *Journal of the Acoustical Society of America*, 104(5):2915–2924, 1998.

A. D. Waite. *Sonar for Practising Engineers*. Thomson Marconi Sonar Limited, 1998.

D. Wang and G. J. Brown. Fundamentals of computational auditory scene analysis. In D. Wang and G.J. Brown, editors, *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*. IEEE Press/Wiley-Interscience, 2006.

D. L. Wang and G. J. Brown. Separation of speech from interfering sounds based on oscillatory correlation. *IEEE Transactions on Neural Networks*, 10(3):684–697, 1999.

K. Wang and S. Shamma. Self-normalization and noise-robustness in early auditory representations. *IEEE Transactions on Speech and Audio Processing*, 2(3): 421–435, 1994.

K. Wang and S. A. Shamma. Zero-crossings and noise suppression in auditory wavelet transforms. Technical Report TR 92-94, University of Maryland, Systems Research Center, 1992.

R. M. Warren. Perceptual restoration of missing speech sounds. *Science*, 167: 392–393, 1970.

M. Weintraub. *A theory and computational model of monaural auditory sound separation*. PhD thesis, Stanford University, 1985.

G. M. Wenz. Acoustic ambient noise in the ocean: Spectra and sources. *Journal of the Acoustical Society of America*, 34(12):1936–1956, 1962.

E. G. Wever. *Theory of hearing*. Wiley, New York, 1949.

A. D. Whalen. *Detection of signals in noise*. Academic Press, New York; London, 1971.

I. C. Whitfield and E. F. Evans. Responses of auditory cortical neurons to stimuli of changing frequency. *Journal of Physiology*, 28:655–672, 1965.

C. C. Wier, W. Jesteadt, and D. M. Green. Frequency discrimination as a function of frequency and sensation level. *Journal of the Acoustical Society of America*, 61(1): 178–184, 1977.

M. C. M. Wright, editor. *Lecture Notes on the Mathematics of Acoustics*. Imperial College Press, 2005.

X. Yang, K. Wang, and S. A. Shamma. Auditory representations of acoustic signals. Technical Report TR 1991-16, University of Maryland, Systems Research Center, 1991.

G. K. Yates. Auditory-nerve spontaneous rates vary predictably with threshold. *Hearing Research*, 57:57–62, 1991.

E. D. Young and M. B. Sachs. Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory nerve fibers. *Journal of the Acoustical Society of America*, 1966(5):1381–1403, 1979.

V. W. Young and P. C. Hines. Perception-based automatic classification of impulsive-source active sonar echoes. *Journal of the Acoustical Society of America*, 122(3):1502–1517, 2007.

G. Zweig, R. Lipes, and J. R. Pierce. The cochlear compromise. *Journal of the Acoustical Society of America*, 59(4):975–982, 1976.

E. Zwicker and R. Feldtkeller. *The Ear as a Communication Receiver*. Acoustical Society of America, 1967/1999. Translated into English by H. Müsch, S. Buus and M. Florentine.

J. J. Zwislocki. Theorie der schneckenmechanik. *Acta Oto-Laryngologica Supplement*, 72, 1948.