

Modelling of Multi-Channel Audio Signals

A dissertation submitted to the **University of Cambridge**
for the degree of **Doctor of Philosophy**

Christopher Martin Hicks

November 18, 1999
Churchill College
Department of Engineering

Declaration

I declare that this dissertation, whose length does not exceed 65,000 words, is purely my own work, and does not represent any work, or the results of any work done in collaboration. It has not been submitted for consideration for any other degree.

CHRISTOPHER HICKS

June 24, 1999

This copy incorporates minor alterations required by the examiners for the award of the degree of Doctor of Philosophy.

November 18, 1999

Acknowledgements

My thanks are due to many people who have helped to make this project possible:

My supervisor, Professor Peter Rayner, has provided many hours of fruitful discussion and plentiful advice, and I thank him for his seemingly endless patience. In addition, Dr Malcolm Macleod and Dr Simon Godsill have given valuable advice and much conversation.

My employer, CEDAR Audio Ltd. has been a source of inspiration and has loaned audio equipment used for preparation of the demonstration CD. CEDAR has been particularly accomodating during the final stages of writing.

Mr E Kendall transcribed recordings from his private archive as test material and for the demonstration CD.

My countless friends at Churchill College, especially Gillian Brown, Rebecca Herisone, Alan Findlay, and the late Sean Cooney have freely given me their time and companionship during times both happy and sad.

My parents have always given me their unwavering love and support.

Kathrin, for everything.

This work was funded by the Engineering and Physical Sciences Research Council.

Summary

THIS DISSERTATION is concerned with the mathematical modelling of musical and audio signals. The emphasis is on multi-channel signals where either more than one copy of a single original is available for analysis, or where the signal comprises two or more parts. The most common example of this latter class is stereo signals which comprise a left and a right signal to create an auditory illusion of space.

Two models are analysed in which we have multiple observations of a single signal. Both are based on the well-known auto-regressive (AR) model which has previously been successfully deployed in many audio applications.

The first of these is the *Multiply-Observed AR Model* in which a single AR signal is contaminated by a number of independent interference signals to give multiple noisy observations of the original. It is shown that the statistics of the noise sources can be determined given certain broad assumptions. The model is applied to the problem of broadband noise reduction of a 78 r.p.m. record, of which a number of copies are available.

The second model is the *Ensemble-AR Model* in which an ensemble of excitation sources drive identical AR filters to give multiple observed signals. Methods for estimation of the AR parameters from the observed data are derived. The model is applied to the detection of impulsive noise in audio signals, and interpolation of the missing data. The E-AR model is demonstrated to be superior to a similar single-channel approach in both of these areas.

There is such a variety of stereo signals in existence that a very general model is needed to encompass their whole spectrum. The *Coupled-ARMA Model* put forward here is based on the ARMA model, but generates a pair of interdependent signals. Its structure

allows efficient estimation of its parameters, and various methods for this are examined. Interpolators for Coupled-ARMA signals are derived.

For much multi-channel audio work it is necessary to ensure that the observed signals are accurately aligned with each other. Where multiple copies of a disc or tape are under examination this is a difficult problem, since even minute time offsets and speed fluctuations lead to effects such as time-varying comb-filtering when the signals are summed. We examine this problem in detail, and develop a robust scheme for resynchronising signals in a Bayesian statistical framework.

Quantisation of audio signals has received much recent research effort. The final part of the dissertation presents a flexible model-based quantisation algorithm. The algorithm is demonstrated in the quantisation of narrow-band signals, and as a powerful enhancement to a simple linear prediction coding system.

Keywords: signal processing — digital audio — signal modelling

Notation

MATHEMATICAL notation used in this dissertation is, for the most part, highly standard. It is as consistent as possible where this is not a hindrance to clarity. The following table indicates the principal typographical styles:

x	scalar
\underline{x}	column vector
$x[i]$ or x_i	element i of \underline{x}
$x(t)$	continuous-time signal
$x[n]$	discrete-time signal
$x[nT]$	discrete-time signal, sample rate $\frac{1}{T}$
X	matrix
$X[i, j]$ or $X_{i,j}$	element i, j of X

In addition, certain symbols are used consistently for particular entities:

x, y	observed data
u, v	hidden data
e, w	white noise source
d, n	disturbance or noise source
a, b	model parameters
σ^2	variance of a random process
N	length of a vector whose index is time
P	length of a model or filter parameter vector
Q	number of channels in a multi-channel system

Furthermore, upper and lower cases of the same letter tend to be closely related.

For example a matrix whose name is X will contain data related to the vector \underline{x} , and the index n may take values $1 \leq n \leq N$.

Contents

1	Introduction	1
1.1	Scope of the Dissertation	2
1.2	Dissertation Overview	3
1.3	Demonstration CD	6
1.4	Colour Figures	6
2	Review	7
2.1	The General Signal Model	9
2.2	Model Parameter Estimation	11
2.3	Single-Channel AR Model	16
2.4	AR Model Parameter Estimation	19
2.5	The ARMA Model	24
2.6	Applications of Audio Signal Models	26
2.7	Audio Signals and the Human Ear	29
2.8	Conclusions	36

3	Multiply-Observed AR (MO-AR) Model	37
3.1	Model Analysis	39
3.2	Signal Estimation	41
3.3	Maximisation of the Conditional Density	47
3.4	Application to Audio Restoration	49
3.5	Conclusions	53
4	Ensemble-AR (E-AR) Model	54
4.1	Ensemble-AR Parameter Estimation	57
4.2	Interpolation of Missing Data	59
4.3	Impulsive Noise Detection	65
4.4	Two-Channel Impulsive Noise Detection	72
4.5	Application to Gramophone Record Restoration	78
4.6	Conclusion	83
5	Coupled ARMA (C-ARMA) Model	84
5.1	Introduction	85
5.2	Application to Stereo Audio Signals	86
5.3	C-ARMA Model Parameter Estimation	88
5.4	Interpolation of Missing Data	106
5.5	Interpolation Tests	112
5.6	Conclusions	113
6	Channel Synchronisation	115
6.1	Introduction	117
6.2	Adaptive Filtering Method	121
6.3	Correlation Method	123
6.4	Model-Enhanced Correlation Method	127
6.5	Statistical Method	129

6.6	Bayesian Formulation	131
6.7	Models and Priors for the Offset	131
6.8	Offset Posterior PDF	137
6.9	Tests of Model-Based Bayesian Estimator	139
6.10	Audio Demonstration	140
6.11	Conclusions	140
7	Model-Based Quantisation	143
7.1	Scalar Quantisation	144
7.2	Model-Based Quantisation	150
7.3	Quantisation of Narrowband Signals	151
7.4	An Enhanced Linear Prediction Coder	157
7.5	Conclusions	164
8	Conclusions	165
8.1	Signal Models	165
8.2	Synchronisation	166
8.3	Applications	166
8.4	Further Research	167
A	Demonstration CD	170
B	Correlation Calculations	172
B.1	Efficient Estimation of the Cross-correlation Function	173
C	Integrals and the Gaussian PDF	176
D	MO-AR Model Error Variances	178
D.1	Weighted Estimate Error Power	178
D.2	Unweighted Estimate Error Power	179

D.3	Comparison of weighted and unweighted signal estimates	180
E	Least Squares and Associated Algorithms	182
E.1	Ordinary Least Squares Method	182
E.2	Approximate Least Squares	182
E.3	Total Least Squares Method	183
E.4	Computational Considerations	185
F	Resampling of a Sampled Signal	186
F.1	Filter Design	187
F.2	Efficient Implementation	188
G	Colour Figures	190
H	References & Bibliography	204

SIGNAL MODELLING is concerned with the mathematical description of data, and as such is a subset of the more general area of data modelling. Data modelling is concerned with the analysis and parameterisation of data for purposes of statistical description, classification, data compression, interpolation, forecasting and so on.

The term “signal” is used to denote a quantity which is related to a physical phenomenon, such as length, luminosity or voltage, and how that quantity varies, frequently as a function of time, but possibly as a function of some other independent variable such as space.

A “time series” is a sequence of data samples, each being associated with a particular instant in time. Thus, if we sample a signal at a number of instants in time then the result is a time series which represents the variation of the physical quantity.

The process of sampling and digitising continuously-varying quantities to form such time series has been well understood for many decades [90, 89]. Refinements continue to be made, particularly in sampling at non-uniform rates, and in the quantisation of signal samples [30]. In particular the quantisation of audio signal samples has received much recent research effort [68, 66, 31].

The field of Digital Signal Processing (DSP) has grown enormously in the last

few decades since it has become feasible to perform complicated calculations on these digitised data sets using general purpose computers and dedicated signal processing systems [47]. Introductions to the practicalities of performing digital signal processing operations on real audio data streams are given in, for example [47, 54].

Audio signal modelling is academically interesting in its own right, but is of an additional interest in an engineering context. There are many engineering applications of signal models in areas such as data compression, restoration, synthesis, and machine interpretation and classification of music, speech and more general audio signals.

1.1 Scope of the Dissertation

The present dissertation is concerned with developing models for high-quality musical signals, and illustrates these with applications. The signal to noise ratios are high (typically 50–80 dB), and the signals may be considered to be stationary over short time-scales of up to a few tens of milliseconds. We concentrate on algorithms and techniques which may be implemented efficiently. It is intended that it will be possible to implement the methods and algorithms presented here in real-time if not currently, then in the readily foreseeable future.

The emphasis of the dissertation is on multi-channel signals, where either multiple copies of a signal are available for analysis, or where a signal has more than one component.

In the former category we include monophonic sources where a number of signals can be extracted from the carrier. For example, many recording media, such as magnetic tape and vinyl records, distribute the stored information along a line, rather than at a point. If this line can be sampled at more than one point then we can extract more than one signal. Such signals all contain the musical information, but any interference from degradation or damage to the medium will be different.

The most important class of signals in the latter category are stereo signals which have two components, the left and the right channels. These signals use differences between the two components to generate the illusion of spatially-separated sources. However, since the two components either originate in the same acoustic space, or are designed to simulate a single acoustic space, they are far from

independent, and it is possible to exploit the correlations to advantage.

Primary applications for these multi-channel models include signal data compression and coding, the area of signal restoration and noise reduction, and signal separation. All of these areas have seen great progress in recent years, but there have been few attempts to use multi-channel techniques in many of these applications.

1.2 Dissertation Overview

Chapter 2

In chapter 2 we review that literature and previous research which is important to a full understanding of subsequent chapters and the new work presented there.

A number of signal models that have been used in audio applications are examined, and we concentrate particularly on the Auto-Regressive, or AR model. This model is central to a large proportion of the literature, and forms the basis of much of the new research presented in this dissertation.

A discussion of applications and justifications for wishing to create signal models is included. The areas of audio restoration and audio coding are covered in some detail, being principal amongst the potential applications for new signal models and analysis techniques.

Also provided is a brief review of sound recording techniques, concentrating particularly on the recording of stereo signals. Much of the justification for the stereo signal model presented in chapter 5 depends upon this material.

Chapters 3 & 4

In chapters 3 and 4 we examine the case where multiple copies of a recording are available. In recognition of the fact that we observe several single-channel signals that convey the same musical information we term these “multiple-mono” systems or signals.

The observed component signals all originate from the same musical information. They may, however, have been degraded and distorted by different mechanisms, or by more than one instance of the same mechanism.

The former case includes situations where a signal has been conveyed by multiple

channels, each having different characteristics, to give distinct observations of that signal. For example, a music recording may be available both as a cassette tape, and as a vinyl LP; the source material—in this case the master from which the LP and cassette were derived—is the same for each, but it is affected differently by the two duplication processes and storage media.

The latter case includes signals that have been conveyed by multiple channels of the same type. The most obvious scenario is that in which multiple copies of, for example, an LP are available. However, in all storage media traditionally used for musical recordings the signal information at a particular instant is spatially distributed, and this allows multiple signals to be read from what appears to be a single source. An simple example of this technique is by use of a stereo pickup to replay a monophonic gramophone record.

In the case where an original, and a transcription of that original (for example, an LP and a cassette tape copy of that LP) it may seem, at first sight, that the copy is of no use to us, since any imperfect transcription represents a loss of information. However, if the cassette in this example had been made some time ago, and the LP has suffered degradation since then, then both are valuable sources. The cassette may well suffer from more broadband noise, but it will lack impulsive noise created by scratches made on the LP subsequent to the transfer.

The linear Auto-Regressive (AR) model has been successfully applied to many areas of single-channel music and speech processing. Chapters 3 and 4 of the dissertation present multi-input multi-output systems, based on the single-channel AR system, which allow analysis and processing multiple-mono signals of the types described above.

Applications in the area of audio restoration that are presented include multi-channel broadband noise reduction, and a two-channel approach to the removal of impulsive noise from monophonic gramophone records. In both cases significant advantages and performance improvements are demonstrated over equivalent single-channel methods.

Chapter 5

Stereophonic signals, which have two components fed to loudspeakers to the left and right of the listener, represent the vast majority of the recorded music archive. They generate a spatial illusion by presenting the left and right ears with non-identical sound pressure waves. This enables virtual sound sources to be placed

anywhere in the horizontal plane between the loudspeakers. Binaural recordings, recorded or synthesised specifically for playback over headphones are successful in placing virtual sources anywhere in three-dimensional space with respect to the listener.

The vast majority of material recorded from around 1960 onwards has been recorded in this format, and it is by far the most common distribution format, being used almost exclusively for all current music distribution media, including CD, MiniDisc and broadcast.

In chapter 5 we present a general model for stereo signals. This system, the C-ARMA model, is shown to be effective for signal estimation and interpolation.

Chapter 6

When multiple copies of a record or tape are available it is usually necessary to ensure that the copies are accurately synchronised with each other, before attempting to process the signals from them together. As well as a time-origin offset there are usually speed fluctuations associated with one or both of the copies, such that the synchronisation of the sources becomes a dynamic problem.

Chapter 6 contains work which provides some insight into this problem, and several algorithms which help achieve this signal synchronisation.

Chapter 7

A novel application of signal models is presented in chapter 7. Quantisation of a signal introduces noise, and we present a model-based quantisation scheme which automatically adapts the power spectrum of this added noise according to the signal characteristics. This is shown to be of benefit when quantising a narrow-band signal. A second application, where it enhances the performance of a Linear Prediction CODEC, is also presented.

Chapter 8

The final chapter contains the conclusions drawn from the whole dissertation, and includes suggestions of areas considered worthy of further research.

1.3 Demonstration CD

A CD of recorded audio examples and demonstrations accompanies the dissertation. This illustrates and demonstrates some of the ideas presented in the text. The tracks on this CD are referenced in the text by their track numbers, *e.g.* [13](#). Appendix A gives a complete track listing of the CD, and cross-references to the main text.

The CD should be playable in any standard CD player. Although formal, controlled listening tests are not included, it is necessary that the audio equipment and listening environment be of a high quality for some of the demonstrations to be effective.

1.4 Colour Figures

Where it is necessary for clarity for figures to be printed in colour they have been moved to appendix G.

2.1	The General Signal Model	9
2.2	Model Parameter Estimation	11
2.2.1	Least Squares Estimation	11
2.2.2	Total Least Squares Estimation	12
2.2.3	Maximum Likelihood Estimation	12
2.2.4	Bayesian Parameter Estimation	14
2.3	Single-Channel AR Model	16
2.3.1	Matrix Representation of the AR Model	17
2.4	AR Model Parameter Estimation	19
2.4.1	Covariance Method	19
2.4.2	Correlation Method	20
2.4.3	Total Least Squares Method	21
2.4.4	Maximum Likelihood Method	22
2.4.5	Bayesian Method	22
2.4.6	Comparison and Conclusions	23
2.5	The ARMA Model	24
2.6	Applications of Audio Signal Models	26
2.6.1	Audio Restoration	26
2.6.2	Audio Coding	27

2.6.3	Audio Signal Synthesis	29
2.7	Audio Signals and the Human Ear	29
2.7.1	Microphones	30
2.7.2	Stereophonic Signals	32
2.7.3	Multi-Channel Audio Signals	35
2.8	Conclusions	36

THE FIRST PART of this review examines previous work in the field of modelling of time-series representations of audio signals, as distinct from other forms of time series. A general and flexible model structure is discussed, and the well-known autoregressive model is examined within this framework. Parameter estimation methods are discussed, with the emphasis being on computationally efficient algorithms.

The second part gives a brief introduction to music and audio signals and recordings. This background is important to understanding the justifications for the models and methods presented later. In particular, the stereo signal models presented in chapter 5 exploit structure and redundancy in the stereo signal which stems from the methods used to create such signals.

2.1 The General Signal Model

A general signal model may be specified by a structure or form, \mathcal{M} , and a set of $P_{\mathcal{M}}$ model parameters, which may be arranged into a column vector $\underline{\mathbf{b}}_{\mathcal{M}}$.

This general model may be concisely represented by the vector equation

$$\underline{\mathbf{x}} = \underline{\mathbf{f}}_{\mathcal{M}}(\underline{\mathbf{b}}_{\mathcal{M}}) + \underline{\mathbf{e}} \quad (2.1)$$

where $\underline{\mathbf{x}}$ is a column vector of N observed data and $\underline{\mathbf{f}}_{\mathcal{M}}(\cdot)$ is a vector-function of

the model parameters $\underline{\mathbf{b}}_{\mathcal{M}}$. This model is almost universally applicable as it is always possible to arrange the observed data and the model parameters each as a column vector.

The model structure is implicit in the length \mathbf{P} of the parameter vector, and the form of function $\underline{f}_{\mathcal{M}}(\cdot)$. It is usual that $\mathbf{N} > \mathbf{P}$ and that the model is therefore a compact representation of the data. This feature forms the basis of many applications, such as coding and classification.

Vector $\underline{\mathbf{e}}$ is interpreted variously, dependent on context and application, as modelling error, observation noise, or as an innovation or excitation sequence. It is also frequently useful to consider the function $\underline{f}_{\mathcal{M}}(\underline{\mathbf{b}}_{\mathcal{M}})$ as a prediction of the data vector $\underline{\mathbf{x}}$, and $\underline{\mathbf{e}}$ as the error associated with this prediction.

Models that are linear in the parameters may be expressed as the linear matrix equation

$$\underline{\mathbf{x}} = \mathbf{F}_{\mathcal{M}} \underline{\mathbf{b}}_{\mathcal{M}} + \underline{\mathbf{e}} \quad (2.2)$$

where $\mathbf{F}_{\mathcal{M}}$ has dimension $\mathbf{N} \times \mathbf{P}_{\mathcal{M}}$. This linear form encompasses two specific model types which are central to the present dissertation. Firstly, $\mathbf{F}_{\mathcal{M}}$ may comprise a set of constant basis vectors (for example complex exponentials), in which case the prediction is simply a weighted sum of those vectors. Secondly, the Auto-Regressive (AR) model may be represented in this framework, in which case both $\underline{\mathbf{x}}$ and $\mathbf{F}_{\mathcal{M}}$ contain observed data. The AR model is analysed in detail in section 2.3.

Notice that the matrix $\mathbf{F}_{\mathcal{M}}$ may contain non-linear functions of known data, such as polynomial or trigonometric functions of the data $\underline{\mathbf{x}}$, without affecting the linearity of the model with respect to its parameters $\underline{\mathbf{b}}_{\mathcal{M}}$.

Throughout the present dissertation we assume the model structure, justifying it from physical principles, the published literature, and experience. The dependence on the structure can therefore become implicit, and the subscript \mathcal{M} will therefore be dropped from here on for notational clarity.

The model equation may therefore be written as

$$\underline{\mathbf{x}} = \underline{f}(\underline{\mathbf{b}}) + \underline{\mathbf{e}} \quad (2.3)$$

$$\underline{\mathbf{x}} = \mathbf{F} \underline{\mathbf{b}} + \underline{\mathbf{e}} \quad (2.4)$$

for the general and linear models respectively.

Choosing the most appropriate model from a candidate set has close ties with the problem of signal classification. The reader is referred to work by, for example, Rajan[81, 82], Duda[23] and Akaike[5] for a deeper treatment of the model selection problem itself, and associated measures of model “fit” such as the AIC [5], BIC [6], MDL [84] and Bayesian evidence [82, 91].

2.2 Model Parameter Estimation

The problem of estimating the model parameters for an assumed model frequently arises. Generally we will have a sample \underline{x} of N observed data, and from these we wish to estimate the P model parameters \underline{b} for an assumed model structure. This will be done either by minimising, in some sense, the vector \underline{e} , or by determining an estimate parameter set according to some underlying statistical model of the data.

2.2.1 Least Squares Estimation

If we assume that the parameters \underline{b} are unknown constants then the least squares (LS) estimate \underline{b}_{LS} is defined as that value of \underline{b} which minimises the sum of the squared errors $\mathcal{E} = \underline{e}^T \underline{e}$ for some observed finite-length \underline{x} .

Since the LS estimate is the one which truly minimises the error energy \mathcal{E} , it is particularly useful in applications which rely on the function $F\underline{b}$ to be an accurate prediction of \underline{x} , Linear Prediction Coding (LPC) being a prime example.

2.2.1.1 General Case

We may rearrange the general model equation 2.3 as

$$\underline{e} = \underline{x} - \underline{f}(\underline{b}) \quad (2.5)$$

and then derive the sum squared error energy $\mathcal{E} = \underline{e}^T \underline{e}$ as

$$\mathcal{E} = (\underline{x} - \underline{f}(\underline{b}))^T (\underline{x} - \underline{f}(\underline{b})). \quad (2.6)$$

To obtain the LS estimate it is required to minimise \mathcal{E} over \underline{b} . In general this will require a complicated non-linear optimisation procedure.

2.2.1.2 Linear Case

In the linear case where $\underline{f}(\underline{b}) = F\underline{b}$ equation 2.6 is quadratic in \underline{b} and has a single minimum which may be found by standard differential calculus.

Differentiating equation 2.6 with respect to the elements of $\underline{\mathbf{b}}$ we obtain

$$\frac{\partial \mathcal{E}}{\partial \underline{\mathbf{b}}} = -2\mathbf{F}^T \underline{\mathbf{x}} + 2\mathbf{F}^T \mathbf{F} \underline{\mathbf{b}}. \quad (2.7)$$

Setting this to zero and solving for $\underline{\mathbf{b}}$ we obtain the location of the turning point

$$\underline{\mathbf{b}}_{[\frac{\partial \mathcal{E}}{\partial \underline{\mathbf{b}}} = 0]} = (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T \underline{\mathbf{x}} \quad (2.8)$$

provided that the inverse exists. It can be shown that this turning point is a global minimum for positive definite $\mathbf{F}^T \mathbf{F}$ and therefore that

$$\underline{\mathbf{b}}_{\text{LS}} = (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T \underline{\mathbf{x}} \quad (2.9)$$

is the value of $\underline{\mathbf{b}}$ which minimises \mathcal{E} .

This method is used at many points throughout the dissertation (sometimes with slight variations), and in most cases the result of this minimisation will be stated without derivation.

2.2.2 Total Least Squares Estimation

The LS estimation procedure implicitly assumes that the function $\underline{f}(\cdot)$ is known. Sometimes however, it will be dependent on noisy experimental observations. In this case, and if the model is linear, the Total Least Squares (TLS) method provides an alternative solution.

The general linear model is rewritten

$$(\mathbf{F} - \mathbf{E}) \underline{\mathbf{b}} = \underline{\mathbf{x}} - \underline{\mathbf{e}} \quad (2.10)$$

where $\underline{\mathbf{e}}$ represents the errors associated with vector $\underline{\mathbf{x}}$, and \mathbf{E} the additional errors associated with \mathbf{F} .

The estimate $\underline{\mathbf{b}}_{\text{TLS}}$ is that value of $\underline{\mathbf{b}}$ which minimises the Frobenius norm¹ of the matrix $[\mathbf{E} \ \underline{\mathbf{e}}]$.

Appendix E describes the TLS method in detail.

2.2.3 Maximum Likelihood Estimation

The LS and TLS methods both treat the elements of the model definition 2.3 as either known or unknown constants, and derive a model estimate as a result of direct algebraic manipulation of the model equation.

¹The Frobenius norm of a matrix \mathbf{A} is defined as the square root of the sum of the squares of its elements $\|\mathbf{A}\|_F = (\sum_{i,j} a_{ij}^2)^{\frac{1}{2}}$.

Alternatively we may treat the observed data \underline{x} as a random variable with a p.d.f., in which case we may define the likelihood function

$$\mathcal{L}(\underline{x}; \underline{b}) = p_{\underline{x}|\underline{b}}(\underline{x} | \underline{b}) \quad (2.11)$$

as the p.d.f. of the data given the true model parameters. The Maximum Likelihood (ML) estimate $\underline{b}_{\text{ML}}$ is defined as the value of \underline{b} which maximises $\mathcal{L}(\underline{x}; \underline{b})$. If we know or assume statistical properties for the error vector \underline{e} then it is usually possible to derive an algebraic expression for $\mathcal{L}(\underline{x}; \underline{b})$ and frequently possible to maximise it analytically.

An important case is where we assume white Gaussian noise of variance σ_e^2 and zero mean for $p_{\underline{e}}(\underline{e})$. Under this condition we may write

$$p_{\underline{e}}(\underline{e}) = \frac{1}{(2\pi\sigma_e^2)^{N/2}} \exp\left(-\frac{\underline{e}^T \underline{e}}{2\sigma_e^2}\right) \quad (2.12)$$

and since $\underline{e} = \underline{x} - \underline{f}(\underline{b})$

$$p_{\underline{e}}(\underline{x} - \underline{f}(\underline{b})) = \frac{1}{(2\pi\sigma_e^2)^{N/2}} \exp\left(-\frac{(\underline{x} - \underline{f}(\underline{b}))^T (\underline{x} - \underline{f}(\underline{b}))}{2\sigma_e^2}\right). \quad (2.13)$$

It can be shown that this p.d.f. is related to the likelihood by the relationship

$$\mathcal{L}(\underline{x}; \underline{b}) = \frac{1}{J(\underline{x}, \underline{b})} p_{\underline{e}}(\underline{x} - \underline{f}(\underline{b})) \quad (2.14)$$

where the Jacobian is defined for the transformation $\underline{x} = \underline{f}(\underline{b})$ as

$$J(\underline{x}, \underline{b}) = \text{abs} \left(\det \left[\frac{\partial \underline{f}^T}{\partial \underline{b}} \right] \right) \quad (2.15)$$

Furthermore, if the prediction $f_n(\underline{b})$ of x_n is linear in x_n itself and does not depend on future elements x_i ($i > n$) then this Jacobian is unity or a simple scale factor and the p.d.f. of equation 2.13 is proportional to the likelihood $\mathcal{L}(\underline{x}; \underline{b})$.

It is often convenient to take the logarithm of 2.13, such that the ‘‘log-likelihood’’ $l(\underline{x}; \underline{b})$ (assuming $J = 1$) is given by

$$l(\underline{x}; \underline{b}) = -\frac{N}{2} \ln(2\pi\sigma_e^2) - \frac{1}{2\sigma_e^2} (\underline{x} - \underline{f}(\underline{b}))^T (\underline{x} - \underline{f}(\underline{b})) \quad (2.16)$$

which may be maximised by any of the standard methods. Since $\ln(\cdot)$ is a monotonically increasing function this maximisation of the log-likelihood yields an identical result to direct maximisation of the likelihood itself.

2.2.3.1 Linear Case

In the linear case the maximisation is once again analytic by differentiation and yields the result

$$\underline{\mathbf{b}}_{\text{ML}} = (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T \underline{\mathbf{x}} \quad (2.17)$$

and the ML estimate is thus seen to be identical to the LS estimate for a finite data sample, under the assumption of a white Gaussian error vector.

2.2.4 Bayesian Parameter Estimation

If we extend the statistical approach further we may treat the model parameters $\underline{\mathbf{b}}$ as random variables, as well as the observed data. In doing so the parameters are assigned a joint p.d.f. $p_{\underline{\mathbf{b}}}(\underline{\mathbf{b}})$ which can reflect either *a-priori* knowledge of the parameters, or merely a degree of uncertainty about them. In the extreme case we can assign a uniform p.d.f. which treats any parameter set as being as likely as any other. Choice of this prior is discussed in more detail later.

Using Bayes' Rule we may express the posterior p.d.f. of the model parameters given the observed data

$$p_{\underline{\mathbf{b}}|\underline{\mathbf{x}}}(\underline{\mathbf{b}} | \underline{\mathbf{x}}) = \frac{p_{\underline{\mathbf{x}}|\underline{\mathbf{b}}}(\underline{\mathbf{x}} | \underline{\mathbf{b}}) p_{\underline{\mathbf{b}}}(\underline{\mathbf{b}})}{p_{\underline{\mathbf{x}}}(\underline{\mathbf{x}})} \quad (2.18)$$

in terms of the likelihood given by equation 2.11, and the prior $p_{\underline{\mathbf{b}}}(\underline{\mathbf{b}})$ which reflects any knowledge we have of the parameters *before* we make the observation $\underline{\mathbf{x}}$, as described above.

The final term, $p_{\underline{\mathbf{x}}}(\underline{\mathbf{x}})$, is known as the *evidence*. It is, in the present context, of little interest as it is constant for any given observation $\underline{\mathbf{x}}$, and hence does not affect the model parameter estimation problem. It does become important in model selection and signal classification problems where it may be calculated as

$$p_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}) = \int_{\underline{\mathbf{b}}} p_{\underline{\mathbf{x}}|\underline{\mathbf{b}}}(\underline{\mathbf{x}} | \underline{\mathbf{b}}) d\underline{\mathbf{b}} \quad (2.19)$$

when required.

Having obtained the posterior distribution there are various options available for sampling a parameter estimate from it. Two convenient estimates which have useful mathematical properties are the minimum mean square error estimate $\underline{\mathbf{b}}_{\text{MMSE}}$ and the maximum *a-posteriori* estimate $\underline{\mathbf{b}}_{\text{MAP}}$.

They correspond to taking the mean

$$\underline{\mathbf{b}}_{\text{MMSE}} = \int_{\underline{\mathbf{b}}} \underline{\mathbf{b}} p_{\underline{\mathbf{b}}|\underline{\mathbf{x}}}(\underline{\mathbf{b}}|\underline{\mathbf{x}}) d\underline{\mathbf{b}} \quad (2.20)$$

and the mode

$$\underline{\mathbf{b}}_{\text{MAP}} = \underset{\underline{\mathbf{b}}}{\operatorname{argmax}} \{p_{\underline{\mathbf{b}}|\underline{\mathbf{x}}}(\underline{\mathbf{b}}|\underline{\mathbf{x}})\} \quad (2.21)$$

respectively, of the posterior distribution. For posterior distributions which are symmetric about the mode (which includes the commonly-encountered Gaussian), the two parameter estimates coincide.

2.2.4.1 Choice of Prior

The seemingly arbitrary choice of prior is the criticism most frequently aimed at proponents of the Bayesian methodology. It is certainly true that choice of a wildly inappropriate prior can give erroneous or misleading results. It is also, however, amongst the most powerful features of the technique, allowing the rigorous and analytically tractable inclusion of even subjective prior information about the problem.

By altering the choice of prior $p_{\underline{\mathbf{b}}}(\underline{\mathbf{b}})$ it is possible to influence the solution to any desired degree. A strong prior will heavily bias the solution; conversely, as the prior becomes flatter compared with the likelihood function, so its influence decreases.

Two commonly-chosen priors, which possess many useful properties, are the Gaussian and uniform distributions. The Gaussian leads to many results being analytically tractable, while providing means to influence the problem to any desired degree by altering the mean and covariance of the distribution. The uniform prior $p(\underline{\mathbf{b}}) = 1$ treats any parameter set as being as likely as any other, and as such imparts no influence upon the solution.

For parameters which form scale-values (for example, the variance σ^2 of a random process) there are other, more appropriate priors. The Gamma distribution, defined for $\alpha, \beta > 0$ as

$$p_G(y|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} \exp(-\beta y) \quad (0 < y < \infty) \quad (2.22)$$

where $\Gamma(\cdot)$ represents the Gamma function itself (see, for example, [60]) is one such example. Appropriate choice of the parameters α and β allows great flexibility

in choosing the degree of influence of the prior. The form of this function yields many analytic results, and in particular the marginalisation of scale parameters with Gaussian likelihoods.

A further much-used example is the improper Jeffreys prior [51]

$$p_J(\mathbf{y}) = \frac{1}{\mathbf{y}} \quad (2.23)$$

It should be noted that this, like the uniform prior, is not normalised to have a unit integral. The Jeffreys prior can be viewed as the limit of the Inverted Gamma distribution

$$p_{IG}(\mathbf{y} \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \mathbf{y}^{-(\alpha+1)} \exp\left(-\frac{\beta}{\mathbf{y}}\right) \quad (0 < \mathbf{y} < \infty) \quad (2.24)$$

as $\alpha \rightarrow 0$ and $\beta \rightarrow 0$.

2.2.4.2 Influence of the Prior

The prior has the effect of biasing the ML parameter estimate towards the value that would be obtained by consideration of the prior alone. The degree to which this occurs depends upon the relative “peakiness” of the likelihood function and prior p.d.f.

There are two important asymptotic conditions, which are independent of the form of prior chosen for $\underline{\mathbf{b}}$. Firstly, as the prior tends to a uniform density (for example, as the variance of a Gaussian prior tends to infinity), the covariance matrix inverse $\mathbf{C}_{\underline{\mathbf{b}}}^{-1} \rightarrow 0$ and the MAP solution tends towards the ML estimate. Secondly, as the number of data points $N \rightarrow \infty$ the likelihood becomes increasingly peaked and once again the solution $\underline{\mathbf{b}}_{\text{MAP}} \rightarrow \underline{\mathbf{b}}_{\text{ML}}$.

2.3 Single-Channel AR Model

The majority of the work represented by the literature in the area of audio signal modelling has been motivated by the desire for machine recognition, coding and synthesis of human speech. However, much of it is applicable to more general audio and musical signals, and in particular the Auto-Regressive (AR) model has been successfully applied to the processing of both speech and music signals.

The AR model has close links with the technique of Linear Prediction, and models

the signal samples as a weighted sum of past samples

$$x[n] = \sum_{i=1}^P a_i x[n-i] + e[n] \quad (2.25)$$

where $e[n]$ is a white Gaussian innovations sequence.

Therrien [97] provides an in-depth analysis of the AR model, and a useful overview is provided by Makhoul [70]. The most important and useful results concerning the AR model and its analysis are summarised here.

The AR finite difference equation 2.25 represents a linear filter whose transfer function is given by

$$\frac{1}{A(z)} = \frac{1}{1 - \sum_{i=1}^P a_i z^{-i}}. \quad (2.26)$$

We see that the AR model represents an all-pole filter (albeit with an order P zero at the origin). Since for the AR model the input to this filter $e[n]$ is white, the signal power spectral density is given by

$$S_{xx}(\omega) = \left| \frac{\sigma_e^2}{1 - \sum_{i=1}^P a_i e^{-j\omega iT}} \right|^2 \quad (2.27)$$

where $1/T$ is the sample rate. The power spectrum shape is therefore determined entirely by the AR model coefficients a_i . The relative phases of the signal components are determined by the innovations sequence $e[n]$, and the power of the signal (for a given set of coefficients) by its variance σ_e^2 .

It is a convenient notation to write expressions such as

$$X(z) = E(z) \frac{1}{A(z)} \quad (2.28)$$

to represent the AR model but it should be borne in mind that since, for the AR model, $e[n]$ is a stochastic process, it is not possible to evaluate its z -transform.

2.3.1 Matrix Representation of the AR Model

The difference equation of the AR model, equation 2.25, may be written in several matrix forms.

2.3.1.1 Direct Form

Consider a block of N contiguous data samples, $x[1 \cdots N]$. If we assemble samples $x[P + 1]$ to $x[N]$ into a column vector \underline{x} then we may write

$$\begin{bmatrix} x[P + 1] \\ x[P + 2] \\ \vdots \\ x[N - 1] \\ x[N] \end{bmatrix} = \begin{bmatrix} x[P] & \cdots & x[1] \\ x[P + 1] & \cdots & x[2] \\ \vdots & & \vdots \\ x[N - 1] & \cdots & x[N - P] \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_P \end{bmatrix} + \begin{bmatrix} e[P + 1] \\ e[P + 2] \\ \vdots \\ e[N - 1] \\ e[N] \end{bmatrix} \quad (2.29)$$

or in a more compact matrix notation

$$\underline{x} = \mathbf{X} \underline{a} + \underline{e}. \quad (2.30)$$

We have chosen to call this the *direct* form, since it is closely related to the representation of the AR model as an IIR filter. This matrix representation of the difference equation is seen to be identical to the linear form of the general model, equation 2.4.

Notice that the first P data samples are contained within matrix \mathbf{X} . If instead we were to prepend the data with P samples $x[-P + 1] \cdots x[0]$ of value zero then we could include all of the observed data in the extended data \underline{x}' thus

$$\begin{bmatrix} x[1] \\ x[2] \\ \vdots \\ x[N - 1] \\ x[N] \end{bmatrix} = \begin{bmatrix} x[0] & \cdots & x[-P + 1] \\ x[1] & \cdots & x[-P + 2] \\ \vdots & & \vdots \\ x[N - 1] & \cdots & x[-P + N] \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_P \end{bmatrix} + \begin{bmatrix} e[1] \\ e[2] \\ \vdots \\ e[N - 1] \\ e[N] \end{bmatrix} \quad (2.31)$$

or in matrix notation

$$\underline{x}' = \mathbf{X}' \underline{a} + \underline{e}'. \quad (2.32)$$

We shall use the direct form in our consideration of parameter estimation techniques.

2.3.1.2 Inverse Form

Alternatively we may rearrange the difference equation

$$e[n] = x[n] - \sum_{i=1}^P a_i x[n - i] \quad (2.33)$$

$$e[n] = \sum_{i=0}^P a'_i x[n] \quad (2.34)$$

where $a'_0 = 1$ and $a'_i = -a_i$, $1 \leq i \leq P$. In this form we have expressed the innovations sequence as an FIR filter applied to the signal $x[n]$. In other words, this arrangement of the model equation represents more closely the *inverse* filter which transforms the observed signal into its associated innovations sequence.

Arranging the samples $e[P + 1]$ to $e[N]$ as a column vector gives the matrix equation

$$\begin{bmatrix} e[P + 1] \\ \vdots \\ e[N] \end{bmatrix} = \begin{bmatrix} -a_p & \cdots & -a_1 & 1 & 0 & \cdots & 0 \\ 0 & -a_p & \cdots & -a_1 & 1 & \cdots & 0 \\ \vdots & & & & & & \vdots \\ 0 & \cdots & 0 & -a_p & \cdots & -a_1 & 1 \end{bmatrix} \begin{bmatrix} x[1] \\ \vdots \\ x[P] \\ x[P + 1] \\ \vdots \\ x[N] \end{bmatrix} \quad (2.35)$$

and if we define $\underline{x}_0 = [x[1] \cdots x[P]]^T$

$$\underline{e} = A \begin{bmatrix} \underline{x}_0 \\ \underline{x} \end{bmatrix}. \quad (2.36)$$

Both of these forms 2.30 and 2.36 will be useful throughout the dissertation.

2.4 AR Model Parameter Estimation

We now turn our attention to the problem of estimating the AR parameters a_i from the observed data \underline{x} .

The covariance and correlation methods are described first. These names appear to have been widely adopted although the terminology is somewhat slack; neither method makes any great distinction between the covariance and autocorrelation functions, and both are generally applied assuming data with zero mean.

2.4.1 Covariance Method

The first method for AR model parameter estimation that we shall examine is known as the ‘‘covariance method’’ [70]. Application of the LS method directly to the matrix AR equation 2.30 gives the result

$$\underline{a}_{LS} = (X^T X)^{-1} X^T \underline{x}. \quad (2.37)$$

The matrix product $M = (X^T X)$, whose elements are of the form

$$M_{ij} = \sum_n x[n-i] x[n-j] \quad (2.38)$$

may be constructed efficiently by noting that the summation for element (i, j) and that for element $(i+1, j+1)$ share all but one of the terms in their respective summations. Furthermore, it is a symmetric matrix $M_{ij} = M_{ji}$, and only one half of it need be calculated directly.

If we view $e[n]$ as the error associated with the prediction of $x[n]$ then the LS parameter estimate has the property that, by definition, it minimises the total prediction error energy over the block of data.

2.4.2 Correlation Method

The correlation method also begins with the matrix form of equation 2.30, and then calculates the excitation energy

$$\mathcal{E} = \underline{e}^T \underline{e} \quad (2.39)$$

$$= \underline{x}^T \underline{x} - 2\underline{x}^T X \underline{a} + 2\underline{a}^T X^T X \underline{a}. \quad (2.40)$$

Whereas the LS method minimises \mathcal{E} directly, the correlation method, by contrast, takes the expectation of this expression and minimises the *expected* value of \mathcal{E} .

$$E[\mathcal{E}] = E[\underline{x}^T \underline{x} - 2\underline{x}^T X \underline{a} + 2\underline{a}^T X^T X \underline{a}] \quad (2.41)$$

$$= (N - P)r_{xx}(0) - 2P\underline{r}_{xx}^T \underline{a} + 2P\underline{a}^T R_{xx} \underline{a} \quad (2.42)$$

where

$$R_{xx} = \begin{bmatrix} r_{xx}(0) & r_{xx}(1) & \cdots & r_{xx}(P-1) \\ r_{xx}(1) & r_{xx}(0) & \cdots & r_{xx}(P-2) \\ \vdots & \vdots & & \vdots \\ r_{xx}(P-1) & r_{xx}(P-2) & \cdots & r_{xx}(0) \end{bmatrix} \quad (2.43)$$

$$\underline{r}_{xx} = \begin{bmatrix} r_{xx}(1) \\ \vdots \\ r_{xx}(P) \end{bmatrix} \quad (2.44)$$

and

$$r_{xx}(i) = E[x[n] x[n-i]]. \quad (2.45)$$

Hence matrix \mathbf{R}_{xx} is the auto-correlation matrix of process $x[n]$ up to lag $P - 1$ and it is this that gives the algorithm its name. Note that we have assumed x to be a stationary process and that \mathbf{R}_{xx} is therefore symmetric.

Minimisation of $E[\mathcal{E}]$ with respect to \underline{a} yields the result

$$\underline{a}_{\text{COR}} = \mathbf{R}_{xx}^{-1} \underline{r}_{xx}. \quad (2.46)$$

The correlation method has several important features.

If the true auto-correlation function and the model order P are known then equation 2.46 gives the true AR coefficients. As such, equation 2.46 represents a fundamental relationship between the AR model parameters and the signal autocorrelation function.

Secondly, the parameters given by equation 2.46 are guaranteed to form a stable filter with all of its poles inside the unit circle. This property stems from the fact that the symmetric Toeplitz matrix formed from the autocorrelation coefficients is guaranteed to be positive definite, and this in turn implies a minimum-phase filter. An outline of the proof is given in [97].

Finally, since matrix \mathbf{R} is Toeplitz, equation 2.46 may be very efficiently solved by Durbin's method [25]. This recursive algorithm provides the parameter estimates in $\mathcal{O}(P^2)$ operations, compared with $\mathcal{O}(P^3)$ for ordinary matrix inversion.

The parameter estimates obtained in any particular case are clearly dependent on the algorithm chosen to estimate the autocorrelation. Since one of the primary motivations for using the correlation method is its computational efficiency, it makes sense to estimate the autocorrelation using an efficient FFT-based method.

The simplest such method calculates the function

$$\hat{\mathbf{R}}_{xx}(m) = \frac{1}{N} \sum_{i=m+1}^N x[i] x[i - m], \quad 0 \leq m < N \quad (2.47)$$

and uses two FFTs, each of length $2N$, and an additional $2N$ complex multiplications. Details of the method are given in appendix B.

2.4.3 Total Least Squares Method

Inspection of equation 2.30 shows that both the matrix \mathbf{X} and the vector \underline{x} contain observed data values, and will both, therefore, be subject to observation noise.

This is precisely the justification that was given for the use of the TLS method in place of the ordinary LS algorithm.

The TLS method provides a parameter estimate $\underline{\mathbf{a}}_{\text{TLS}}$ which will, in some cases, be a better estimate of the true model than that given by the ordinary LS algorithm. The TLS method is, however, highly computationally expensive, requiring the calculation of the SVD of an $(N - P)$ by $(P + 1)$ matrix.

Experience has shown that the TLS algorithm has a tendency to place poles on or outside the unit circle when the data is very noisy, or when the data set is relatively small.

2.4.4 Maximum Likelihood Method

The likelihood $\mathcal{L}(\underline{\mathbf{x}}; \underline{\mathbf{a}})$ for the AR model under the assumption of white Gaussian excitation $\mathbf{e}[\mathbf{n}]$ may be obtained by substituting terms into 2.14 to obtain

$$\mathcal{L}(\underline{\mathbf{x}}; \underline{\mathbf{a}}) = \frac{1}{(2\pi\sigma_e^2)^{(N-P)/2}} \exp\left(-\frac{(\underline{\mathbf{x}} - \mathbf{X}\underline{\mathbf{a}})^\top(\underline{\mathbf{x}} - \mathbf{X}\underline{\mathbf{a}})}{2\sigma_e^2}\right). \quad (2.48)$$

It was shown above that, under these conditions, the ML parameter estimate coincides with the LS estimate. That is

$$\underline{\mathbf{a}}_{\text{ML}} = \underline{\mathbf{a}}_{\text{LS}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \underline{\mathbf{x}}. \quad (2.49)$$

It should be noted that $\underline{\mathbf{x}}$ comprises samples $[\mathbf{x}[P + 1] \cdots \mathbf{x}[N]]^\top$ and that there is an implicit conditionality on the P initial samples $\underline{\mathbf{x}}_0 = [\mathbf{x}[1] \cdots \mathbf{x}[P]]^\top$. Thus the likelihood should be more properly written as $\mathcal{L}(\underline{\mathbf{x}}; \underline{\mathbf{a}}, \underline{\mathbf{x}}_0)$.

The discrepancies between the parameter estimates given by maximisation of $\mathcal{L}(\underline{\mathbf{x}}; \underline{\mathbf{a}}, \underline{\mathbf{x}}_0)$ as against maximisation of the true likelihood $\mathcal{L}([\underline{\mathbf{x}}_0 \ \underline{\mathbf{x}}]^\top; \underline{\mathbf{a}})$ are small if $N \gg P$. Godsill [39, 35] and Box *et al.* [11] give alternative derivations of an expression for this exact likelihood should that be required.

2.4.5 Bayesian Method

Recall that in the Bayesian framework we treat the model parameters as random variables, and then sample their distribution to give parameter estimates with properties suitable for a specific application.

We have previously given in equation 2.18 the posterior p.d.f. of the model parameters for the general model, given the observed data. Furthermore, we

have given in equation 2.48 the likelihood function $\mathcal{L}(\underline{x}; \underline{a}) = p_{\underline{x}|\underline{a}}(\underline{x} | \underline{a})$ for the AR model.

Substituting the likelihood into the expression for the posterior density we obtain

$$p_{\underline{a}|\underline{x}}(\underline{a} | \underline{x}) = \frac{1}{(2\pi\sigma_e^2)^{(N-P)/2}} \exp\left(-\frac{(\underline{x} - \mathbf{X}\underline{a})^\top(\underline{x} - \mathbf{X}\underline{a})}{2\sigma_e^2}\right) \frac{p_{\underline{a}}(\underline{a})}{p_{\underline{x}}(\underline{x})} \quad (2.50)$$

and having chosen a suitable prior $p_{\underline{a}}(\underline{a})$ we may maximise this directly to obtain $\underline{a}_{\text{MAP}}$. Recall that the evidence term, $p_{\underline{x}}(\underline{x})$ is constant over \underline{a} .

The general multi-variate Gaussian

$$p_{\underline{a}}(\underline{a}) = \frac{1}{(2\pi|C_{\underline{a}}|)^{P/2}} \exp\left(-\frac{1}{2}(\underline{a} - \underline{m}_{\underline{a}})^\top C_{\underline{a}}^{-1}(\underline{a} - \underline{m}_{\underline{a}})\right) \quad (2.51)$$

with mean $\underline{m}_{\underline{a}}$ and covariance $C_{\underline{a}}$ is a convenient prior since it leads to an analytic maximisation. Substituting 2.51 into equation 2.50 and taking the logarithm gives the log posterior density. Maximising with respect to \underline{a} yields the MAP parameter estimate

$$\underline{a}_{\text{MAP}} = (\mathbf{X}^\top \mathbf{X} + \sigma_e^2 C_{\underline{a}}^{-1})^{-1} (\mathbf{X}^\top \underline{x} + \sigma_e^2 C_{\underline{a}}^{-1} \underline{m}_{\underline{a}}). \quad (2.52)$$

We see in this expression that the MAP estimate is based upon the ML solution we saw earlier, but now it has been “moulded” by the prior on \underline{a} . As the covariance of the prior increases, the inverse covariance matrix tends to zero, and the influence of the prior on the solution decreases, such that the MAP solution tends to the ML solution. Similarly, as the data set is enlarged, the influence of the data terms $\mathbf{X}^\top \mathbf{X}$ and $\mathbf{X}^\top \underline{x}$ is relatively increased, and again the MAP solution tends to the ML.

2.4.6 Comparison and Conclusions

We have seen that there is a number of methods available for estimation of model parameters. The number of operations required for the different methods varies widely. To demonstrate this, MATLAB was used to count the floating point operations (FLOPS) required to calculate estimates of an order-25 model from 1000 data points, this being typical of the problem size in audio signal processing.

The results are given in table 2.1. The final column shows the approximate proportion of the computational capacity of an inexpensive modern DSP chip [47, 54] that would be required to perform this calculation in real time, assuming the standard professional audio sampling rate of 44.1 kHz.

The correlation method is clearly the cheapest, owing largely to the fact that the Toeplitz system may be solved using Durbin's method. The Least Squares method is possible to realise for this example; being $\mathcal{O}(P^3)$ it would be considerably more practical for lower model orders. The Total Least Squares method is significantly too computationally expensive to be economically feasible in most applications of this type. The SVD involved is $\mathcal{O}((P+1)N^2 + N^3)$ and so is dominated by the data vector length, rather than the model order.

Note that the MAP estimate, and others that may be obtained using the Bayesian method, are omitted from the table. If the solution is analytic with Gaussian prior (equation 2.52) then it is of similar complexity to the LS and ML methods; however, if a less convenient prior were necessary (for sound statistical reasons) then a complicated and expensive optimisation may be required which would increase the FLOP count dramatically.

The more elaborate methods, when simplified sufficiently to allow analytic or efficient numerical solution, yield parameter estimates which are similar, if not identical, to those obtained using simpler methods. Therefore, the algorithms presented in later chapters will, for the most part, use those simpler methods. However, it should be borne in mind that there is always the option of using the more elaborate schemes in specific scenarios where the simpler methods are found to be lacking.

2.5 The ARMA Model

Addition of moving average (MA) terms to the AR model gives the time-domain difference equation

$$x[n] = \sum_{i=1}^P a_i x[n-i] + \sum_{i=0}^P b_i e[n-i] \quad (2.53)$$

Method	FLOPS	% DSP
COR	1.6×10^5	6.9%
LS/ML	1.3×10^6	56%
TLS	1.1×10^8	4700%

TABLE 2.1: *Computational Load to Estimate AR Parameters*

where the additional parameters \mathbf{b}_i define the moving average filter. The excitation signal $e[n]$ is a white Gaussian signal of unit variance. This is known as the ARMA model.

It is possible to express the ARMA difference equation 2.53 in matrix form

$$\underline{\mathbf{x}} = \mathbf{X}\underline{\mathbf{a}} + \mathbf{E}\underline{\mathbf{b}} \quad (2.54)$$

for a finite block of data, analogous to equation 2.30 for the AR model.

This form was used for the estimation of AR model parameters. However, the estimation of ARMA model parameters does not have a unimodal solution analogous to that for the AR model. This fact makes the ARMA model significantly less suitable for real-time applications where computational simplicity is a requirement. Numerous techniques for the parameter estimation problem have been suggested; Priestley [78] and Therrien [98] are useful starting points.

We can also write the equivalent of the matrix inverse form, equation 2.36. This requires the definition of the internal AR process

$$u[n] = \sum_{i=1}^P a_i u[n-i] + e[n] \quad (2.55)$$

such that

$$x[n] = \sum_{i=0}^P b_i u[n-i]. \quad (2.56)$$

These equations may be written in matrix form

$$\underline{\mathbf{e}} = \mathbf{A}\underline{\mathbf{u}} \quad (2.57)$$

$$\underline{\mathbf{x}} = \mathbf{B}\underline{\mathbf{u}} \quad (2.58)$$

where \mathbf{A} and \mathbf{B} are both $\mathbf{N} \times (\mathbf{N} + \mathbf{P})$, vectors $\underline{\mathbf{e}}$ and $\underline{\mathbf{x}}$ are length \mathbf{N} , and $\underline{\mathbf{u}}$ is length $\mathbf{N} + \mathbf{P}$.

The z -domain transfer function of the model is given by

$$\frac{B(z)}{A(z)} = \frac{\sum_{i=0}^P b_i z^{-i}}{1 - \sum_{i=1}^P a_i z^{-i}} \quad (2.59)$$

from which it can be seen that the moving average terms add \mathbf{P} zeros to the signal model.

Given that the excitation signal is white, the signal power spectral density is given by

$$S_{xx}(\omega) = \left| \frac{\sum_{i=0}^P b_i e^{-j\omega i T}}{1 - \sum_{i=1}^P a_i e^{-j\omega i T}} \right|^2 \quad (2.60)$$

It can be shown that the p.s.d. of an AR signal may be made to match that of an ARMA signal to arbitrary accuracy by sufficiently increasing the order of the AR model. The result is that we can safely assume an AR model, provided that we are prepared to allow its order to be relatively large. Since highly efficient algorithms exist for the AR parameter estimation problem this is an attractive approach.

In chapter 5 we present a two-channel signal model which includes moving-average terms, but whose structure allows the parameters to be estimated efficiently. This allows us to exploit the more compact parameterisation of the ARMA model without the overhead of a lengthy parameter estimation.

2.6 Applications of Audio Signal Models

Signal modelling techniques have a broad range of applications across the field of signal processing [16]. It allows convenient extraction and analysis of the form of the data, and as such provides a useful framework for problems of estimation, classification and so on. Signal modelling techniques have been successfully applied in many diverse areas such as seismology, medicine (*e.g.* [7]) and motion-picture restoration [56, 57].

For audio signals, the principal application areas in which models have been successfully employed are signal restoration [39] and noise reduction [39, 63, 64, 62], the areas of signal coding and data compression [12, 19, 58], and signal synthesis [88, 85].

2.6.1 Audio Restoration

Audio restoration is the process of estimating an audio signal from a noisy or corrupted observation of that signal. Audio signals may be stored on analogue discs or magnetic tape, and these media are prone to physical damage and defects (such as scratches) which degrade the audio signal. Real-time communication channels, such as analogue radio and cable links are prone to interference which

similarly may degrade the audio signals they carry.

Model-based methods have been shown in the past (*e.g.* [101, 35, 92, 63, 102]) to be highly effective in this application. Of particular interest is the application of the AR model to the detection and removal of impulsive noise, such as is introduced by scratches on a gramophone record [101, 35]. Details of these methods are given in chapter 4, where new extensions to multi-channel systems are also described. Other models such as the wavelet basis [100] and sinusoidal model [69] have also been successfully applied in this area, as well as methods based on a DFT decomposition of the audio data [17].

2.6.2 Audio Coding

Audio coding is concerned with the compact description of audio data. Frequently the most convenient form in which to manipulate audio data is linear PCM, but this is not a compact form in which to store or transmit it. Audio coding schemes exploit structure in the data to reduce this storage requirement. Signal modelling is a convenient framework within which to analyse and exploit this structure.

Audio coding algorithms fall into two principal categories:

- lossless algorithms (*e.g.* [19, 15, 71]), in which the original PCM data may be reconstructed precisely from the coded data, and
- lossy algorithms (*e.g.* [53, 52, 13, 12]), in which psychoacoustic phenomena are exploited to allow audibly imperceptible data to be eliminated from the coded signal.

In the second of these cases the reconstructed PCM data is not identical to the original, but stimulates the human auditory system in a similar manner. These lossy algorithms are capable of high compression ratios in applications where regeneration of the auditory stimulus is the only requirement.

2.6.2.1 Model-based Audio Coding

Conventional coding and data compression algorithms such as run-length coding and Huffman coding do not work well on audio data. Signal models, however, provide a basis for a class of compression schemes which exploit the structure inherent in audio data to provide a much greater coding gain.

It was shown above that the terms of the model equation 2.1 may be regarded as

a prediction of the signal, and the error associated with this prediction. In this context the model structure and parameters provide an approximate representation of the data, which encapsulates much of its general form. These parameters, combined with a coded form (possibly linear PCM) of the prediction error signal, form the basis of a compact representation of the original audio data.

Details of the algorithm, and variation of the coding of the parameters and error signal give rise to a wide range of audio coders, both lossless and lossy. The signal model chosen is frequently linear for simplicity, and in this case the technique is known as Linear Predictive Coding (LPC). This structure is adaptable to a broad range of applications from the very low bit rate coding of speech to high quality data compression of musical signals.

In chapter 7 we present a new extension to a simple LPC coder which improves its performance when applied to high-quality audio signals.

2.6.2.2 *Multi-Channel Audio Coding*

Many of the audio coding schemes currently in use apparently allow the joint coding of multiple audio channels. The coding algorithms do not, however, include sophisticated methods for exploiting inter-channel redundancy [13, 12], but simply choose, on a frame-by-frame basis whether to code the left and right signals of a stereo pair, or whether to code their sum and difference.

Fuchs presents a scheme [28] for inter-channel prediction within the framework of a sub-band system such as MPEG. The paper shows results for a scheme which predicts the signal in a given sub-band from the signal in the same sub-band in the partnering channel. The predictor is a gross time delay of up to ± 50 samples, plus an order 3 FIR filter. Since the filter is of such a low order it represents principally a delay of sub-sample resolution, together with a little general shaping of the frequency response of the sub-band filter.

A recent algorithm [71] extends the sum/difference model (which can be viewed as a 45 degree rotation of the stereo field) for up to 64 audio channels by allowing the coded streams to be an optimal rotation of the input channels. At the time of writing, this algorithm is the subject of commercial licensing negotiations and authoritative details are therefore difficult to obtain.

2.6.3 Audio Signal Synthesis

We have seen that we can treat the signal model as a highly parameterised representation of an audio signal. It is often the case that these parameters have a tangible relationship with the perceived characteristics of the signal, such as its pitch or timbre. It is possible, then, to alter these parameters, or to excite the model with a synthetic excitation sequence, and thereby synthesise a new signal, which despite being entirely synthetic, retains qualities of the original.

For example, if we have an AR model for a musical tone then we may synthesise a new, similar tone by exciting the AR filter with a suitable synthetic excitation sequence. Furthermore, we create a similar tone of a different pitch by scaling the frequency axis of the power spectral density (recall that this is simply related to the AR model coefficients by equation 2.27), and then suitably exciting this new filter.

Example techniques and applications are given in [88, 85, 65, 4, 21].

2.7 Audio Signals and the Human Ear

The human ear is a complex detector of acoustic signals (*i.e.* pressure waves in air). The approximate frequency range over which the ear operates usefully is 20 Hz to 20 kHz, and it has a dynamic range of approximately 120 dB. It is not uniformly sensitive, and these ranges vary significantly between individuals. A valuable reference for the workings of the ear and the human auditory system is given by Moore [73].

Transducers (microphones and loudspeakers) are available to convert acoustic signals to and from an electrical analogue. The electrical form may be recorded by converting it to a physical form, such as magnetisation on a tape, or modulations of a groove on a gramophone disc². Furthermore, an electrical signal from a microphone, or retrieved from a recording, may be sampled (and quantised) and subsequently stored and processed in a digital form. It is also possible to generate synthetic signals (by means of electronics or a computer algorithm), and convert them to sound with a loudspeaker.

Signals in any of these acoustic, electrical, physical or digital forms, that are

²The earliest recordings were made by using the acoustic signal energy to cut a groove, often in a wax substrate, by purely mechanical means.

destined ultimately for the human ear, are referred to as *audio* signals.

2.7.1 Microphones

The microphone is a device for converting an acoustic signal into an electrical analogue. Many technologies exist to perform this function [1], the principal ones using electromagnetic or electrostatic effects.

Microphones are characterised by their frequency response, and how that response varies with the angle of incidence of the acoustic wave (the *polar pattern*). The microphones used for music and speech recording are usually first-order designs. That is, their polar response pattern (for a distant source) is given by

$$H(\theta) = \alpha + (1 - \alpha) \cos \theta \quad (2.61)$$

where α controls the pickup pattern, and θ is the incident angle of the pressure wave. The restriction that the source be distant ensures that the incident wave is effectively a plane-wave. Microphone designs with fixed polar pattern are most common, but some elaborate designs allow the user to alter α as required.

Four common patterns are shown in figure 2.1, though it should be noted that the names corresponding to the particular values of α are not standardised, and variations are often encountered. Each pattern can be considered as the weighted sum of an omnidirectional microphone which measures pressure, and a figure-8 microphone which measures velocity. Note that the rear lobe (which appears for $\alpha < \frac{1}{2}$) is of opposite polarity to the front lobe. Thus turning a figure-8 microphone through 180° results in a polarity inversion of the signal from it.

If a pressure-sensing (omnidirectional) microphone and a velocity-sensing (figure-eight) microphone are mounted in close proximity then the weighted sum of their outputs results in a signal that effectively comes from a virtual microphone at the effective centre of the pair. The polar pattern of this microphone may be set arbitrarily between the extremes of omnidirectional and figure-8 simply by changing the weights.

The ultimate extension of this idea is the *Soundfield* microphone [33] which, in concept at least, has a pressure output, and three mutually-orthogonal velocity outputs. This set of four signals gives a complete description of the soundfield at the acoustic centre of the microphone. A suitable weighted sum of the four signals effects a virtual microphone of arbitrary (first-order) pattern, pointing arbitrarily in three dimensions. By generating several such sums, any number of

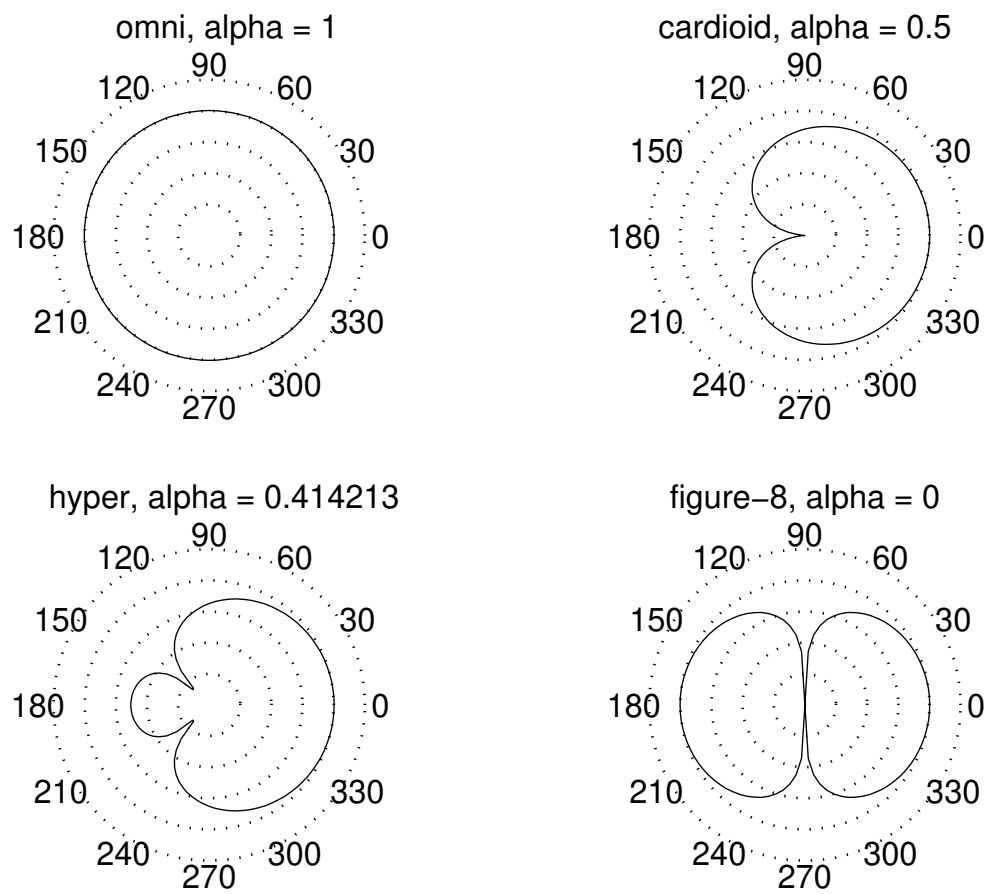


FIGURE 2.1: *The four most common microphone polar patterns, omnidirectional, cardioid, hyper-cardioid and figure-of-eight. The radial scale is in dB, with 10 dB per division. The microphone is in each case nominally pointing to the right.*

coincident microphones may be simulated together.

Further information about the design and engineering of microphones used for high quality audio may be found in [1] and [29].

2.7.2 Stereophonic Signals

The vast majority of music recordings made today are stereo, or two-channel recordings. Two channels are used to generate an illusion of spatial separation between different signal sources, and also an illusion of the acoustic space surrounding those sources. There is frequently some redundancy between the two channels which can be exploited in a signal-modelling scheme, and it this area which is explored in chapter 5.

Stereo recording was pioneered by Blumlein in the 1930's [10] (reprinted in [26]), and since then a number of techniques for making spatially-illusory recordings of this type have been developed. Stereo recordings have been widespread since the mid 1950s [26, 72]; a comprehensive treatment of the subject is given in [26].

The classes of stereo signal which we will consider here are:

Phase Stereo, where the stereo illusion is brought about by the difference in path length from the source to each of a pair of omnidirectional microphones,

Intensity Stereo, where intensity differences between the two channels create the stereo illusion,

Hybrids of intensity and phase, that use a microphone arrangement which records both intensity for each channel, and phase difference information,

Binaural recordings, in which small pressure-measuring microphones are placed in the ear canals of a real or dummy head, so as to capture direction-of-arrival information in the same way as the human head and *pinnae*, and

Synthetic Stereo, in which the left and right signals are generated electronically, and not by microphones in an acoustic space.

A compact reference to these and many other microphone arrangements used for stereo recording is provided by [93].

2.7.2.1 Phase Stereo

Phase stereo is captured by a pair of spaced omnidirectional microphones. The spatial illusion is brought about through the time-of-arrival differences caused by unequal path lengths from the source to the two microphones. This mimics the fact that the ears are separated by several inches, though microphones are often spaced further apart than this.

For a microphone separation of $2d$ and a distant source at angle θ , the signals at the microphone outputs are

$$x_L(t) = x(t + \tau) \quad (2.62)$$

$$x_R(t) = x(t - \tau), \quad (2.63)$$

where τ is given by

$$\tau = \frac{d}{c} \sin \theta, \quad (2.64)$$

and c is the speed of sound in air (figure 2.2).

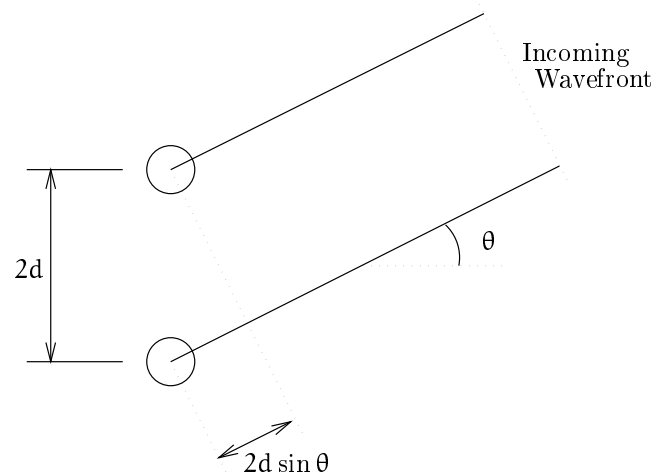


FIGURE 2.2: *Spaced Omnidirectional Microphones*

2.7.2.2 Intensity Stereo

The second class of stereo signals includes those in which the illusion of space is brought about by amplitude differences in the two channels. Such signals are created by a pair of directional microphones which are spatially coincident.

This mimics, to a degree, the shading effect of the head over one ear, of sounds originating from the opposite side.

The most famous intensity technique described by Blumlein [10] is that commonly known to recording engineers as *Blumlein* or *X-Y*. A pair of figure-8 microphones ($\alpha = 0$) are mounted at right angles, and as close together as possible. The resulting axis of symmetry points towards the major sound source.

More generally, other types of microphones may be used, and with different angular spacings. The microphone signals are then ideally given by

$$x_L(t) = \sum_i x_i(t)(\alpha + (1 - \alpha) \cos(\theta_i - \theta_0)) \quad (2.65)$$

$$x_R(t) = \sum_i x_i(t)(\alpha + (1 - \alpha) \cos(\theta_i + \theta_0)). \quad (2.66)$$

where α determines the polar response of the microphones, and θ_0 is half the angular separation between them.

2.7.2.3 Hybrid Techniques

A large number of stereo signals are generated by microphone techniques that draw on a combination of both phase and intensity illusions. An overview is given in [26]; particularly interesting examples are the sphere microphone [96], the Jecklin Disc [50], the Faulkner array [27] and ORTF [93].

It is the prevalence of these types of signals that is the primary motivation for wishing to devise a generalised model for stereo signals. In general a signal source will appear in both channels, but with a differing amplitude and phase in each. These differences occur as a result of the differing incidence angles and unequal path lengths to the microphones respectively, as discussed previously.

2.7.2.4 Binaural Stereo

Arguably giving the most realistic psycho-acoustic illusion, binaural recordings are only really useful where headphones are employed for replay. This is because they rely on the signal from the left microphone reaching just the left ear, and similarly for the right; any cross-talk destroys the illusion.

Binaural recordings are most prevalent in multi-media and virtual-reality applications. There has, in recent years, been significant effort made towards identifying the transfer functions associated with the head and *pinnae* in order to simulate binaural signals without the use of a dummy head.

2.7.2.5 *Synthetic Stereo*

Much use is made of electronics and DSP to enhance musical signals during recording, particularly of “pop” music (which tends to be recorded one part at a time on a multi-track tape machine). Localisation of the sources in the stereo spread is traditionally done with a *pan-pot*. This device splits a single signal in some proportion, and routes each part to one of the channels. This system essentially simulates the intensity stereo described above; there is no phase or frequency response modification of the signal.

Also classed as synthetic stereo are those recordings which are derived from large numbers of “spot” microphones spread around an acoustic space. The signals from each is generally placed in a realistic position in the stereo spread with a pan-pot.

Increasingly, advanced DSP techniques are being used to introduce phase cues, either in addition to, or instead of the intensity cues generated by the pan-pot. The use of this synthetic phase information gives an enhanced impression of “space”, and can even allow sources to be made to appear outside the angle subtended by the loudspeakers.

2.7.3 **Multi-Channel Audio Signals**

There has, in recent years, been an increasing interest in multi-channel audio, particularly in the production of audio for films. This has led to the development of a number of multi-channel audio coding schemes, such as MPEG, AC3 and DTS.

They all incorporate a multi-channel audio track, which typically comprises five full-bandwidth channels (left, right, centre, left rear, right rear) and an additional low frequency, low bandwidth “sub-woofer” channel. The formats are hence frequently referred to as “5.1 channel” schemes.

The multi-channel coding used for these audio formats is relatively simple. The channels are typically treated as independent audio streams, and channel bandwidth is allocated to each from a common pool according to a psycho-acoustic model. Exploitation of redundancy in the signals is limited; details are difficult to obtain as many of these schemes are commercial secrets, but MPEG, for example, allows the sum and difference of a stereo pair to be coded, instead of the left and right components themselves [13].

The most well-known of these psycho-acoustic phenomena is that of tonal masking [73]; a strong tone masks a weaker tone at a nearby frequency. In addition these multi-channel coders make use of spatial masking. A sound source at a particular position has a greater masking effect over a second source at a similar position than over a second source that is separated from it by some angular displacement relative to the listener. In addition, multi-channel sound is often accompanied by pictures which are very suggestive at drawing the listener's attention to predominant sound sources.

2.8 Conclusions

We have seen that signal modelling provides a framework for parameterisation of an audio signal. Various methods for estimating the parameters of a model of assumed structure have been described. The Auto-Regressive (AR) model has been described in detail, and algorithms for determination of its parameters have been compared.

The nature of recorded sound signals has been described, with particular emphasis on stereo signals, which format represents the vast majority of the recorded sound archive. The present emergence of systems that convey more than two discrete channels has been noted, and also the fact that current coding standards for signals of this type make little use of possible inter-channel redundancy.

Multiply-Observed AR (MO-AR) Model

3.1	Model Analysis	39
3.1.1	Conditional PDF of the True Signal	39
3.1.2	Signal Likelihood Function	41
3.2	Signal Estimation	41
3.2.1	Maximum Likelihood Signal Estimation	42
3.2.2	Noise Estimation — Two-Channel	43
3.2.3	Noise Estimation — Multi-Channel	44
3.2.4	Verification of the Signal Estimation Algorithm	45
3.3	Maximisation of the Conditional Density	47
3.3.1	Model Parameter Estimation	48
3.4	Application to Audio Restoration	49
3.4.1	Audio Demonstration	50
3.5	Conclusions	53

Multiply-Observed AR (MO-AR) Model

A SYSTEM in which a single signal is corrupted by a number of interference signals can give multiple observations of the same underlying signal. The system is shown as a block-diagram in figure 3.1, with the underlying signal modelled as autoregressive.

The system outputs are the Q observed signals x_q . A white excitation signal drives an all-pole filter $1/A(z)$ to give the true signal u . This unobservable signal is contaminated by Q noise sources to give the observations $x_q = u + n_q$.

We wish to analyse the observable signals, and from them derive estimates of

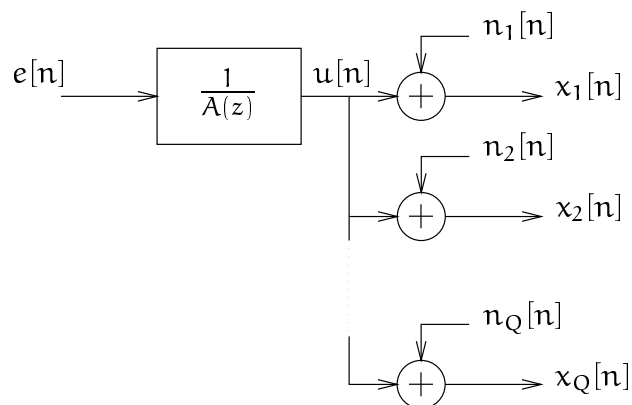


FIGURE 3.1: *Multiple Observations of an AR Process*

the true signal and the model parameters. First we derive expressions for the p.d.f. of the true signal samples. We then show how, in two distinct cases, some statistical properties of the interference sources may be estimated. These two are then combined to give an estimate of the true signal, dependent on the observed data alone.

3.1 Model Analysis

3.1.1 Conditional PDF of the True Signal

Assume, for the moment, that the model parameters $\theta = \{\underline{a}^T, \sigma_e^2\}$ are known. The true AR signal $u[n]$ is given by the expression

$$u[n] = \sum_{p=1}^P a_p u[n-p] + e[n] \quad (3.1)$$

$$u[n] = \hat{u}[n] + e[n] \quad (3.2)$$

where $\hat{u}[n] = u[n] - e[n]$ can be considered an estimate of the signal, and $e[n]$ is a white, Gaussian random variable of variance σ_e^2 .

Since $e[n]$ is drawn from a random process $N(0, \sigma_e^2)$ it is clear from equation 3.2 that the p.d.f. of $u[n]$ given the model parameters is given by

$$p_{u|\theta}(u[n] | \theta) = N(\hat{u}[n], \sigma_e^2), \quad (3.3)$$

assuming an implicit conditionality on $\underline{u} = [u[n-1] \cdots u[n-P]]^T$, the initial condition vector.

Further, the p.d.f. of the observation $x_q[n]$ given $u[n]$ and $\sigma_{n_q}^2$, and assuming Gaussian (though not necessarily white) noise sources, is

$$p_{x|u}(x_q[n] | u[n], \sigma_{n_q}^2) = N(u[n], \sigma_{n_q}^2), \quad (3.4)$$

where $\sigma_{n_q}^2$ is the variance of the q^{th} noise source. Note that this is a scalar equation, including just one sample from the noise source, and hence does not include the noise source covariance matrix.

If the noise sources are independent, then the joint p.d.f. of the observations is the product of the Q individual p.d.f.'s:

$$p_{\underline{x}|u}(\underline{x}_Q[n] | u[n], \underline{\sigma}_n^2) = \prod_{q=1}^Q p_{x|u}(x_q[n] | u[n], \sigma_{n_q}^2), \quad (3.5)$$

where

$$\underline{x}_Q[\mathbf{n}] = [x_1[\mathbf{n}] \cdots x_Q[\mathbf{n}]]^T, \quad (3.6)$$

$$\underline{\sigma}_n^2 = [\sigma_{n_1}^2 \cdots \sigma_{n_Q}^2]^T. \quad (3.7)$$

Bayes' Theorem states that

$$p(\underline{\alpha} | \underline{\beta}, \gamma) = \frac{p(\underline{\beta} | \underline{\alpha}, \gamma) \cdot p(\underline{\alpha} | \gamma)}{p(\underline{\beta} | \gamma)} \quad (3.8)$$

and in this context we treat $\underline{\alpha}$ as the hidden data we wish to estimate, $\underline{\beta}$ as the noisy observations of that data, and γ as a set of model parameters.

We can use Bayes' Theorem to combine equations 3.3 and 3.5 to give

$$p_{u|\underline{x}}(u[\mathbf{n}] | \underline{x}_Q[\mathbf{n}], \underline{\sigma}_n^2, \theta) = \frac{p_{\underline{x}|u}(\underline{x}_Q[\mathbf{n}] | u[\mathbf{n}], \underline{\sigma}_n^2) \cdot p_{u|\theta}(u[\mathbf{n}] | \underline{\sigma}_n^2, \theta)}{p_{\underline{x}}(\underline{x}_Q[\mathbf{n}] | \underline{\sigma}_n^2, \theta)}, \quad (3.9)$$

which expresses the p.d.f. of the true data given the observations as a function of the p.d.f. of the observations given the data (equation 3.5), and the p.d.f. of the data given the model (equation 3.3).

For convenience we define $\phi(u[\mathbf{n}])$

$$\phi(u[\mathbf{n}]) = p_{u|\underline{x}}(u[\mathbf{n}] | \underline{x}_Q[\mathbf{n}], \underline{\sigma}_n^2, \theta) \quad (3.10)$$

as given by equation 3.9.

The denominator of equation 3.9 is constant over variations in $u[\mathbf{n}]$, and may therefore be replaced by a constant of proportionality K . Furthermore, $u[\mathbf{n}]$ has no dependence on $\underline{\sigma}_n^2$. Substituting terms into equation 3.9 therefore gives

$$\phi(u[\mathbf{n}]) = K \left[\prod_{q=1}^Q p_{x|u}(x_q[\mathbf{n}] | u[\mathbf{n}], \sigma_{n_q}^2) \right] \cdot p_{u|\theta}(u[\mathbf{n}] | \theta) \quad (3.11)$$

$$= K \left[\prod_{q=1}^Q \frac{1}{\sqrt{2\pi\sigma_{n_q}^2}} \exp\left(\frac{-n_q^2[\mathbf{n}]}{2\sigma_{n_q}^2}\right) \right] \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left(\frac{-e^2[\mathbf{n}]}{2\sigma_e^2}\right) \quad (3.12)$$

and if we define

$$\Sigma^2 = 2\pi\sigma_e^2 \prod_{q=1}^Q (2\pi\sigma_{n_q}^2) \quad (3.13)$$

then we may simplify further, giving

$$\phi(u[\mathbf{n}]) = \frac{K}{\Sigma} \exp\left(\sum_{q=1}^Q \frac{-n_q^2[\mathbf{n}]}{2\sigma_{n_q}^2} - \frac{e^2[\mathbf{n}]}{2\sigma_e^2}\right) \quad (3.14)$$

Substituting for $e[n] = u[n] - \underline{a}^T \underline{u}$ and $n_q[n] = x_q[n] - u[n]$ gives the p.d.f. of the signal sample $u[n]$ conditional on the observed data, the noise variances, and the AR model parameters as

$$\phi(u[n]) = \frac{K}{\Sigma} \exp \left(\sum_{q=1}^Q \frac{-(x_q[n] - u[n])^2}{2\sigma_{n_q}^2} - \frac{(u[n] - \underline{a}^T \underline{u})^2}{2\sigma_e^2} \right) \quad (3.15)$$

where

$$\underline{a} = [a_1 \dots a_p]^T \quad (3.16)$$

$$\underline{u} = [u[n-1] \dots u[n-P]]^T \quad (3.17)$$

Note that there is an implicit conditionality on the initial condition vector \underline{u} .

3.1.2 Signal Likelihood Function

By consideration of the signals alone, and disregarding for the time being their origins in a common AR model, we may derive the likelihood function for the signal sample $u[n]$ given the multiple observations.

The p.d.f. of $u[n]$ given $x_q[n]$ is straightforwardly given by

$$p_u(u[n] | x_q[n]) \propto \exp \left(-\frac{(u[n] - x_q[n])^2}{2\sigma_{n_q}^2} \right) \quad (3.18)$$

The noise sources are assumed to be independent, so the p.d.f.

$$\phi_m(u[n]) = p_u(u[n] | \underline{x}[n]) \quad (3.19)$$

for $u[n]$ given the Q observed samples $\underline{x}[n] = [x_1[n] \dots x_Q[n]]$ is given by the product

$$\phi_m(u[n]) = \prod_{q=1}^Q p_u(u[n] | x_q[n]) \quad (3.20)$$

$$= K \exp \left(\sum_{q=1}^Q -\frac{(u[n] - x_q[n])^2}{2\sigma_{n_q}^2} \right) \quad (3.21)$$

where K is a normalising constant such that $\int_{-\infty}^{+\infty} \phi_m du$ is unity.

3.2 Signal Estimation

It has been shown in section 3.1.2 that the likelihood function $\phi_m(u[n])$ can be analytically derived, using no assumptions other than Gaussianity and independence

of the interfering signals. We can use this p.d.f., along with knowledge of the noise source variances, to estimate the underlying signal.

3.2.1 Maximum Likelihood Signal Estimation

Differentiation of the likelihood function $\phi_{\mathbf{m}}(\mathbf{u}[\mathbf{n}])$ in order to maximise it with respect to $\mathbf{u}[\mathbf{n}]$ yields the signal estimate

$$\hat{\mathbf{u}}_{\mathbf{x}}[\mathbf{n}] = \left(\sum_{q=1}^Q \frac{1}{\sigma_{n_q}^2} \right)^{-1} \sum_{q=1}^Q \frac{x_q[\mathbf{n}]}{\sigma_{n_q}^2} \quad (3.22)$$

which is a weighted sum of the observations only, with the weights inversely proportional to the noise source variances $\sigma_{n_q}^2$. It can be seen from equation 3.22 that as the signal-to-noise ratio of a particular channel decreases, so does its contribution to the signal estimate.

It is straightforward to show that $\hat{\mathbf{u}}_{\mathbf{x}}[\mathbf{n}]$ is unbiased by making the observation that the expected estimation error $\mathbb{E}[\mathbf{u}[\mathbf{n}] - \hat{\mathbf{u}}[\mathbf{n}]] = \mathbf{0}$. Further, it can be shown¹ that the estimation error variance is given by

$$\mathbb{E} [(\mathbf{u}[\mathbf{n}] - \hat{\mathbf{u}}_{\mathbf{x}}[\mathbf{n}])^2] = \left(\sum_{q=1}^Q \frac{1}{\sigma_{n_q}^2} \right)^{-1} \quad (3.23)$$

and hence the signal to noise ratio of the estimated signal is given by

$$\text{SNR}_{\hat{\mathbf{u}}} = 10 \log_{10} \left(\sigma_{\mathbf{u}}^2 \sum_{q=1}^Q \frac{1}{\sigma_{n_q}^2} \right) \quad (3.24)$$

By comparison, taking the unweighted mean of the observations

$$\bar{\mathbf{x}}[\mathbf{n}] = \frac{1}{Q} \sum_{q=1}^Q x_q[\mathbf{n}] \quad (3.25)$$

as the signal estimate gives an estimation error variance of

$$\mathbb{E} [(\mathbf{u}[\mathbf{n}] - \bar{\mathbf{x}}[\mathbf{n}])^2] = \frac{1}{Q^2} \sum_{q=1}^Q \sigma_{n_q}^2. \quad (3.26)$$

¹Derivations of equations 3.23 and 3.26, and a proof of equation 3.28 all appear in appendix D.

and corresponding signal to noise ratio

$$\text{SNR}_{\bar{x}} = 10 \log_{10} \left(\sigma_u^2 \left(\frac{1}{Q^2} \sum_{q=1}^Q \sigma_{n_q}^2 \right)^{-1} \right) \quad (3.27)$$

The signal to noise ratio $\text{SNR}_{\hat{u}}$ is guaranteed to be at least as high as $\text{SNR}_{\bar{x}}$ since

$$10 \log_{10} \left(\sigma_u^2 \sum_{q=1}^Q \frac{1}{\sigma_{n_q}^2} \right) \geq 10 \log_{10} \left(\sigma_u^2 \left(\frac{1}{Q^2} \sum_{q=1}^Q \sigma_{n_q}^2 \right)^{-1} \right) \quad (3.28)$$

with the case of equality being when all the noise variances are equal.

3.2.2 Noise Estimation — Two-Channel

In the two-channel case the observed signals are given by

$$x_1[n] = u[n] + n_1[n] \quad (3.29)$$

$$x_2[n] = u[n] + n_2[n] \quad (3.30)$$

and we assume that the noise sources n_1 , n_2 are independent but identically distributed (i.i.d.). This is a good model for the continuous broadband noise inherent to many recording media used for music and speech (such as vinyl and magnetic tape) when replayed with a two-channel pickup or head. In this case we may use the difference between the two observed signals

$$d[n] = x_1[n] - x_2[n] \quad (3.31)$$

$$= n_1[n] - n_2[n] \quad (3.32)$$

to estimate the noise distribution.

First we calculate the autocorrelation of this difference signal

$$R_{dd}(m) = E \left[d[n] d[n-m] \right] \quad (3.33)$$

$$= E \left[(n_1[n] - n_2[n]) (n_1[n-m] - n_2[n-m]) \right] \quad (3.34)$$

and since n_1 and n_2 are independent it follows that

$$R_{dd}(m) = E \left[n_1[n] n_1[n-m] \right] + E \left[n_2[n] n_2[n-m] \right] \quad (3.35)$$

$$= R_{n_1 n_1}(m) + R_{n_2 n_2}(m). \quad (3.36)$$

Furthermore, since the statistical properties of n_1 and n_2 are identical

$$R_{nn}(m) = \frac{1}{2} R_{dd}(m) \quad (3.37)$$

where $R_{nn}(\mathbf{m})$ is the noise source autocorrelation function.

Thus the noise autocorrelation function $R_{nn}(\mathbf{m})$ and its power spectral density

$$S_{nn}(\omega) = \sum_{\mathbf{m}=-\infty}^{\infty} R_{nn}(\mathbf{m}) \exp(-j\omega\mathbf{m}T) \quad (3.38)$$

may be estimated from the observed signals x_1 and x_2 .

3.2.3 Noise Estimation — Multi-Channel

In the multi-channel case where $Q \geq 3$ we may extend the two-channel analysis, and thereby remove the requirement that the noise sources be identically distributed. We continue to assume independence of the noise sources.

Consider two of the signals x_i, x_j in isolation, and define their difference to be

$$d_{ij}[\mathbf{n}] = x_i[\mathbf{n}] - x_j[\mathbf{n}]. \quad (3.39)$$

For Q observed signals there are QC_2 possible difference relationships of this form for which $i \neq j$.

From equation 3.36 (note that up to this point we have assumed only independence of the noise sources) we may write

$$R_{d_{ij}d_{ij}}(\mathbf{m}) = R_{n_i n_i}(\mathbf{m}) + R_{n_j n_j}(\mathbf{m}) \quad (3.40)$$

which relates the autocorrelation of the difference signal d_{ij} to the autocorrelations of the two noise sources n_i and n_j .

In systems of three or more channels (*i.e.* if $Q \geq 3$) we can calculate an estimate of the noise autocorrelations by constructing the following matrix equation (shown, for example, for $Q = 4$), which encapsulates equation 3.40 for all $i \neq j$.

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} R_{n_1 n_1}(\mathbf{m}) \\ R_{n_2 n_2}(\mathbf{m}) \\ R_{n_3 n_3}(\mathbf{m}) \\ R_{n_4 n_4}(\mathbf{m}) \end{bmatrix} = \begin{bmatrix} R_{d_{12}d_{12}}(\mathbf{m}) \\ R_{d_{13}d_{13}}(\mathbf{m}) \\ R_{d_{14}d_{14}}(\mathbf{m}) \\ R_{d_{23}d_{23}}(\mathbf{m}) \\ R_{d_{24}d_{24}}(\mathbf{m}) \\ R_{d_{34}d_{34}}(\mathbf{m}) \end{bmatrix} \quad (3.41)$$

Equation 3.41 is solvable in a least-squares sense when $Q \geq 3$ to give

$$\begin{bmatrix} R_{n_1 n_1}(\mathbf{m}) \\ R_{n_2 n_2}(\mathbf{m}) \\ R_{n_3 n_3}(\mathbf{m}) \\ R_{n_4 n_4}(\mathbf{m}) \end{bmatrix} = (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \begin{bmatrix} R_{d_{12} d_{12}}(\mathbf{m}) \\ R_{d_{13} d_{13}}(\mathbf{m}) \\ R_{d_{14} d_{14}}(\mathbf{m}) \\ R_{d_{23} d_{23}}(\mathbf{m}) \\ R_{d_{24} d_{24}}(\mathbf{m}) \\ R_{d_{34} d_{34}}(\mathbf{m}) \end{bmatrix} \quad (3.42)$$

which is the estimate of the autocorrelations at lag \mathbf{m} as required. Matrix \mathbf{M} is the “combination” matrix and has ${}^Q C_2$ rows. The condition $Q \geq 3$ is equivalent to the condition that $\mathbf{M}^T \mathbf{M}$ be non-singular. This matrix is of simple structure, having the value $Q-1$ on the leading diagonal, and 1 elsewhere. It is Toeplitz, and we may therefore solve the system efficiently by Durbin’s method [25], although this is unlikely to be necessary for systems with few channels.

Once the autocorrelation functions have been estimated, the power spectral densities follow from equation 3.38.

3.2.4 Verification of the Signal Estimation Algorithm

Consider an eight-channel system.

A block of synthetic simulation data $\mathbf{u}[\mathbf{n}]$, $1 \leq n \leq 2000$, was generated using a resonant AR(10) process. This data was corrupted with eight independent Gaussian AR(1) signals

$$\mathbf{n}_q[\mathbf{n}] = \sigma_{n_q} (1 - \alpha_q^2) (\mathbf{w}_q[\mathbf{n}] + \alpha_q \cdot \mathbf{n}_q[\mathbf{n} - 1]) \quad (q = 1 \cdots 8) \quad (3.43)$$

to generate eight observed signals

$$\mathbf{x}_q[\mathbf{n}] = \mathbf{u}[\mathbf{n}] + \mathbf{n}_q[\mathbf{n}] \quad (3.44)$$

each of 2000 samples. Signal \mathbf{w}_q is a white Gaussian source of unit variance.

3.2.4.1 Experiment One

The first experiment checks the match between the theoretical SNR of the estimated signal with that which is achieved in practice.

In the eight-channel system described above the SNR of channels 1–7 was held constant. The SNR of channel 8 was swept from -20 dB to +20 dB relative to this. Two estimates of the original signal were made from the corrupted data.

The first, \bar{x} was the unweighted mean of simultaneous samples (equation 3.25), and the second, $\hat{u}_{\underline{x}}$ the weighted sum given by equation 3.22.

The results are plotted in figure 3.2, which shows the recovered signal noise power against the noise power of channel eight. The match between the theoretical curves (solid lines) and the simulation data (crosses) is good.

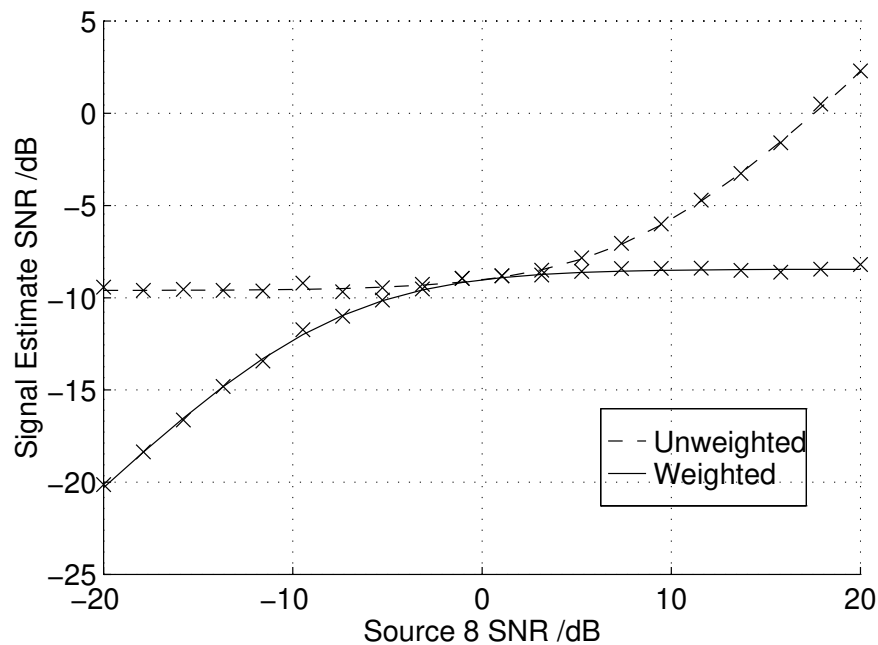


FIGURE 3.2: *Noise Power of Signal Estimates*

The MAP signal estimate $\hat{u}_{\underline{x}}$ is seen to be at least as good as the mean \bar{x} , and much better where there are extreme SNR differences between the channels:

- When the noise power of x_8 is relatively low, the MAP signal estimate is dominated by this signal x_8 .
- When the noise power of x_8 is relatively high, the MAP signal estimate rejects this noisy signal, and bases the signal estimate on x_1 to x_7 .

Where the signals all have equal noise powers the two signal estimates coincide.

3.2.4.2 Experiment Two

Experiment two fixes the SNR of all channels to a range of values spanning approximately 20 dB, and verifies that the improvement noted with the MAP

estimate is consistent. The set of parameters $\{\sigma_{n_q}^2, \alpha_q\}$ of the noise sources chosen for experiment two is given in table 3.1.

q	$\sigma_{n_q}^2$	α_q	SNR/dB
1	0.01	0.90	53.5
2	0.04	0.93	47.5
3	0.09	0.92	44.0
4	0.16	0.85	41.5

q	$\sigma_{n_q}^2$	α_q	SNR/dB
5	0.25	0.88	39.5
6	0.36	0.95	38.0
7	0.49	0.98	36.6
8	0.64	0.97	35.5

TABLE 3.1: *Experimental Parameters*

The noise variances were estimated *via* equation 3.42 with $m = 0$, and these were subsequently used to estimate the underlying data *via* equation 3.22.

Over one thousand trials \bar{x} was found to be, on average, 2.4 dB more noisy than the quietest of $x_1 \dots x_8$, whereas \hat{u}_x was 2.3 dB quieter. In other words, with an approximate 20 dB difference between the observed SNR extremes, the MAP estimate was an average of 4.7 dB better than the unweighted estimate. The MAP estimate was better than the unweighted mean for every one of the 1000 blocks of trial data.

3.3 Maximisation of the Conditional Density

As an alternative to marginalisation, $\phi(u[n])$ may be maximised directly to give the true ML estimate based on the conditional density,

$$\hat{u}_{x,\mathcal{M}}[n] = \left(\sum_{q=1}^Q \frac{x_q[n]}{\sigma_{n_q}^2} + \frac{\underline{a}^T \underline{u}}{\sigma_e^2} \right) \left(\sum_{q=1}^Q \frac{1}{\sigma_{n_q}^2} + \frac{1}{\sigma_e^2} \right)^{-1}. \quad (3.45)$$

This has the same form as equation 3.22 and treats the estimate of $u[n]$ derived from the model as a further observation with noise variance σ_e^2 .

The estimation error variance is therefore given by

$$E [(u[n] - \hat{u}_{x,\mathcal{M}}[n])^2] = \left(\sum_{q=1}^Q \frac{1}{\sigma_{n_q}^2} + \frac{1}{\sigma_e^2} \right)^{-1}. \quad (3.46)$$

Thus, if the model excitation variance is small compared with the smallest additive noise variance, then $\hat{u}_{x,\mathcal{M}}[n]$ will be a better estimate of the signal than $\hat{u}_x[n]$. In order for this to be useful we need to know the model parameters.

3.3.1 Model Parameter Estimation

To estimate \underline{a} it is convenient to consider the data in blocks of length N . We can write down two p.d.f.'s, the block equivalents of equations 3.3 and 3.4 respectively.

$$p_{\underline{u}|\underline{a}}(\underline{u} | \underline{a}, \sigma_e^2) = (2\pi\sigma_e^2)^{-\frac{N-p}{2}} \exp\left(-\frac{\underline{u}^T \mathbf{A}^T \mathbf{A} \underline{u}}{2\sigma_e^2}\right) \quad (3.47)$$

$$p_{\underline{x}_q|\underline{u}}(\underline{x}_q | \underline{u}, \mathbf{R}_{n_q}) = ((2\pi)^N |\mathbf{R}_{n_q}|)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} (\underline{x}_q - \underline{u})^T \mathbf{R}_{n_q}^{-1} (\underline{x}_q - \underline{u})\right) \quad (3.48)$$

where the column vectors \underline{x}_q and \underline{u} are each of length N . \mathbf{R}_{n_q} is the correlation matrix for the q^{th} noise source, and \mathbf{A} is the matrix

$$\mathbf{A} = \begin{bmatrix} -a_p & -a_{p-1} & \cdots & -a_1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -a_p & -a_{p-1} & \cdots & -a_1 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & & \ddots & \ddots & & \vdots \\ 0 & \cdots & 0 & -a_p & -a_{p-1} & \cdots & -a_1 & 1 & 0 \\ 0 & 0 & \cdots & 0 & -a_p & -a_{p-1} & \cdots & -a_1 & 1 \end{bmatrix}. \quad (3.49)$$

These equations represent the p.d.f. of a finite block of true data \underline{u} given the model parameters, and the p.d.f. of a block of observed data \underline{x}_q given the true data \underline{u} and the noise correlation matrix \mathbf{R}_{n_q} .

Using the probability chain rule and the assumption that the noise sources are independent we may write the p.d.f. of *all* the observed data given \underline{u} and the noise correlations as

$$p_{\mathbf{X}|\underline{u}}(\mathbf{X} | \underline{u}, \mathbf{R}_n) = \prod_{q=1}^Q p_{\underline{x}_q|\underline{u}}(\underline{x}_q | \underline{u}, \mathbf{R}_{n_q}) \quad (3.50)$$

$$= \prod_{q=1}^Q ((2\pi)^N |\mathbf{R}_{n_q}|)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} (\underline{x}_q - \underline{u})^T \mathbf{R}_{n_q}^{-1} (\underline{x}_q - \underline{u})\right) \quad (3.51)$$

where \mathbf{X} is the complete set of observed data

$$\mathbf{X} = \begin{bmatrix} \vdots & & \vdots \\ \underline{x}_1 & \cdots & \underline{x}_Q \\ \vdots & & \vdots \end{bmatrix} \quad (3.52)$$

and \mathbf{R}_n represents the correlations of all the noise processes $\mathbf{R}_{n_1} \cdots \mathbf{R}_{n_Q}$.

Once again we can use Bayes' Theorem and the probability chain rule to give

$$p_{\underline{a}|\mathbf{X}}(\underline{a} | \mathbf{X}, \sigma_e^2, \mathbf{R}_n) = \frac{p_{\mathbf{X}|\underline{u}}(\mathbf{X} | \underline{u}, \mathbf{R}_n) p_{\underline{u}|\underline{a}}(\underline{u} | \underline{a}, \sigma_e^2) p_{\underline{a}}(\underline{a}, \sigma_e^2) p_{\mathbf{R}}(\mathbf{R}_n)}{p_{\mathbf{X}}(\mathbf{X})} \quad (3.53)$$

where $p_{\underline{a}}(\underline{a}, \sigma_e^2)$ and $p_{\mathbf{R}}(\mathbf{R}_n)$ represent any *a-priori* knowledge of the model and noise parameters that we may have. Since $p_{\mathbf{X}}(\mathbf{X})$ is constant over \underline{a} we may write

$$p_{\underline{a}|\mathbf{X}}(\underline{a} | \mathbf{X}, \sigma_e^2, \mathbf{R}_n) \propto p_{\mathbf{X}|\underline{u}}(\mathbf{X} | \underline{u}, \mathbf{R}_n) p_{\underline{u}|\underline{a}}(\underline{u} | \underline{a}, \sigma_e^2) p_{\underline{a}}(\underline{a}, \sigma_e^2) p_{\mathbf{R}}(\mathbf{R}_n). \quad (3.54)$$

The MAP parameter estimate $\underline{a}_{\text{MAP}}$ is given by

$$\underline{a}_{\text{MAP}} = \underset{\underline{a}}{\operatorname{argmax}} \{ p_{\mathbf{X}|\underline{u}}(\mathbf{X} | \underline{u}, \mathbf{R}_n) p_{\underline{u}|\underline{a}}(\underline{u} | \underline{a}, \sigma_e^2) p_{\underline{a}}(\underline{a}, \sigma_e^2) p_{\mathbf{R}}(\mathbf{R}_n) \} \quad (3.55)$$

whose two component likelihood functions are given by equations 3.51, and 3.48, and where $p_{\underline{a}}(\underline{a}, \sigma_e^2)$ and $p_{\mathbf{R}}(\mathbf{R}_n)$ are Bayesian priors on the model parameters and noise correlations respectively.

Equation 3.55 represents a difficult optimisation problem and its full solution is outside the scope of the present work. There has been much study of high-dimensionality probability density functions of this type, and it is likely that, for example, Monte Carlo Markov Chain and Gibbs' sampling methods [34] would be applicable to the present problem. These methods have been successfully applied to associated audio signal problems by a number of researchers [75, 76, 99, 41].

We have shown previously that \mathbf{R}_n may be estimated by independent means, and these estimates may be incorporated as strong Bayesian priors. Sampling methods are highly computationally expensive, owing primarily to their iterative nature; incorporating such priors is expected to be of great benefit in speeding the convergence to a solution, particularly where there is a large number of these parameters in a multi-channel system.

3.4 Application to Audio Restoration

Multiple copies of musical recordings are frequently available, and in most cases the noise sources that contaminate each are approximately independent. For example, if a microphone signal were recorded simultaneously to two tapes, then the noise inherent to the recording medium (the tape hiss) is independent for the two. This practice of making a simultaneous backup has been common since the earliest days of recording. In this case it is clear that the algorithms presented in this chapter are applicable.

If only a single copy of a recording is available it may be possible to extract multiple signals from it by use of, for example, a multi-track tape head. Recording studios regularly record two tracks onto tape half an inch wide. Off-the-shelf

tape heads can be bought today which have sixteen tracks across this width, thus enabling eight copies of each signal to be extracted.

3.4.1 Audio Demonstration

Four copies of an archive disc were available for study. A two-channel transcription of each provided a total of eight signals from which to prepare a restored copy.

3.4.1.1 Pre-processing

These raw two-channel transcriptions can be heard on tracks [1]–[4] of the demonstration CD. The impulsive noise was removed from each of these transcriptions with the commercial CEDAR audio restoration system [18]. They were then synchronised using techniques from chapter 6.

3.4.1.2 Signal Restoration — Spectral Subtraction

Firstly, each track was restored using the spectral subtraction method [37], based on automatically-estimated noise spectra. The noise spectrum contaminating each signal was estimated by averaging equation 3.42 (on an on-going basis) over several seconds (and hence over several revolutions of the disc).

Figure 3.3 shows a short excerpt of the signal from the two groove walls of [2]. The spectrum of the outer wall signal is shown as the upper part of figure 3.4, and the estimated spectrum of the noise content is shown below it.

The multi-channel nature of this system allows continual update of the estimated noise spectra through the course of the extract. This is not possible in the case of a single-channel system because signal components contaminate the noise estimate. These single-channel systems typically require a noise estimate to be made from an otherwise silent part of the track, and then assume the noise to be of constant spectral density throughout the extract.

Following the spectral subtraction algorithm the eight resulting signals were averaged, and this average signal is presented as track [5].

3.4.1.3 Signal Restoration — Statistical Method

Secondly, the signal was estimated using equation 3.22, and this restoration is presented on track [6]. We would expect this restoration not to be so good as the

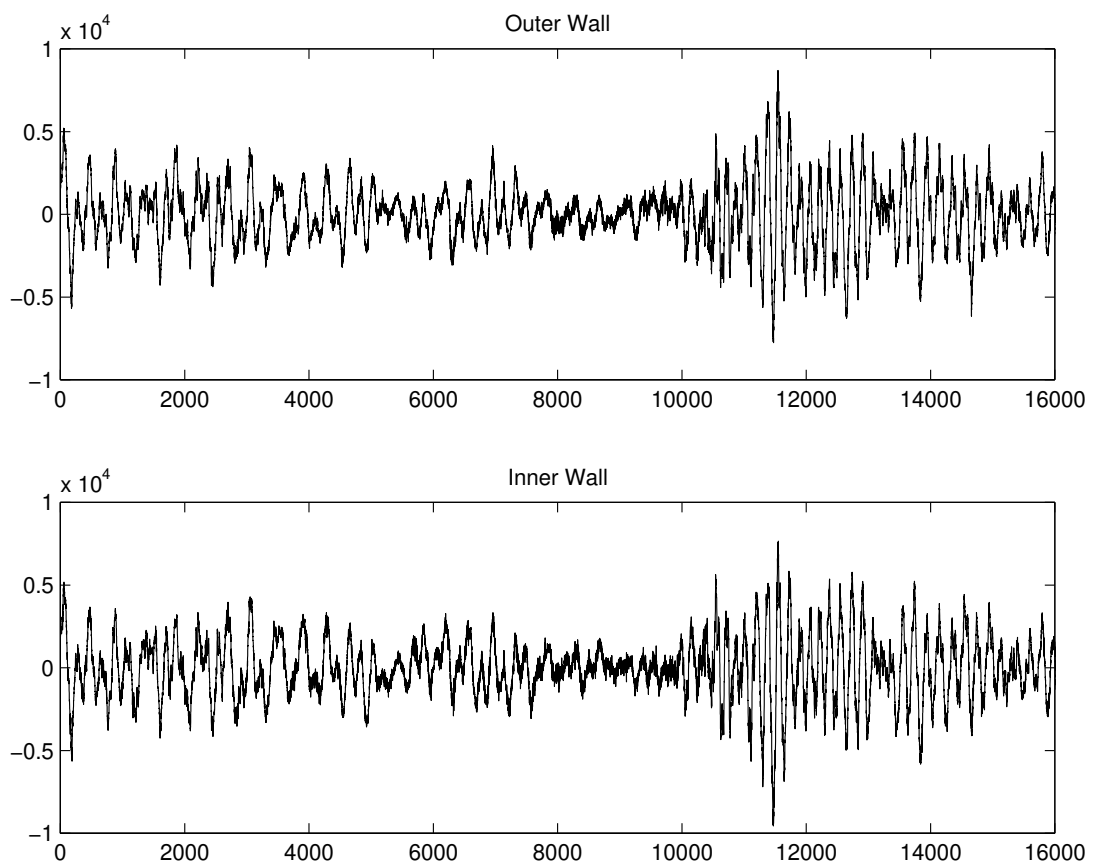


FIGURE 3.3: *Signals from 78 r.p.m. gramophone record.*

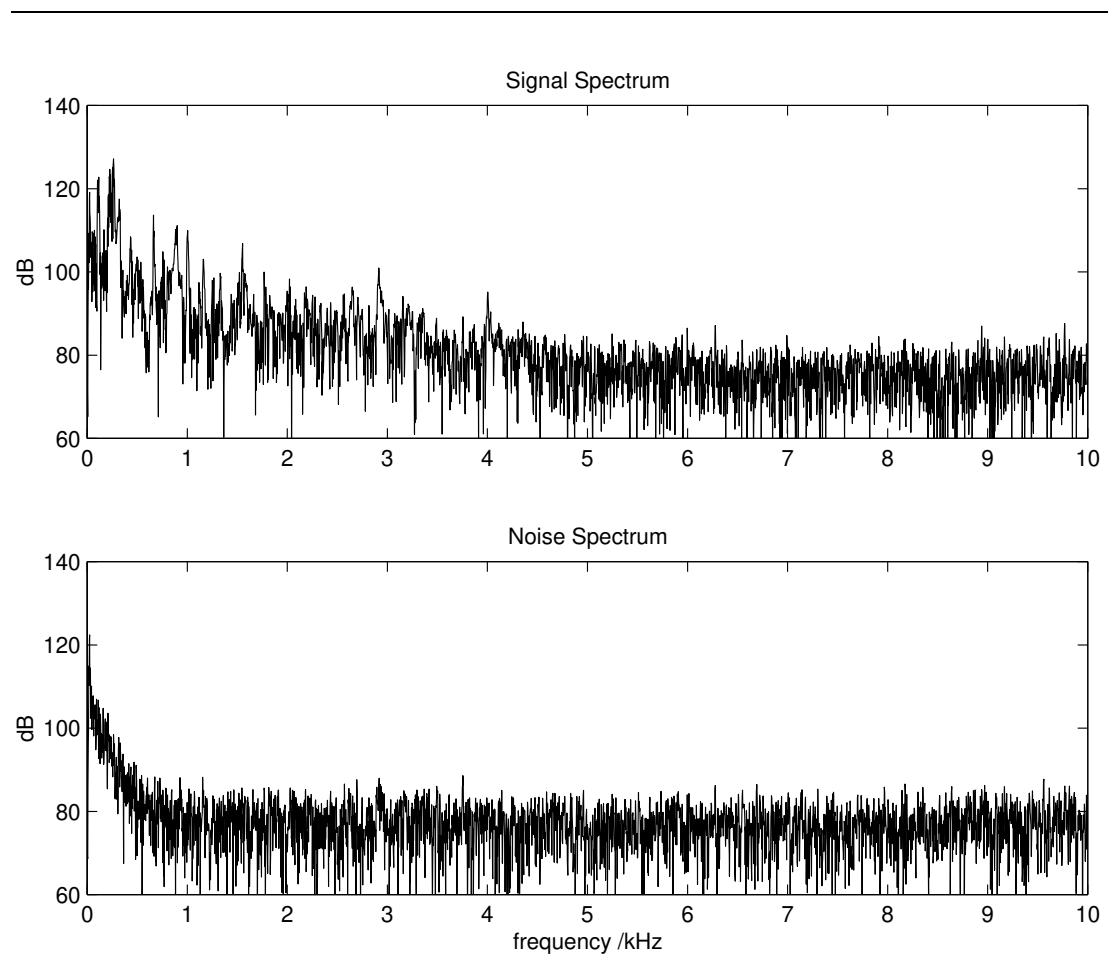


FIGURE 3.4: *Estimated signal and noise spectra from outer groove wall.*

spectral subtraction method since it does not take into account the colours of the interfering signals, and this is found to be the case. It is, however, a useful noise reduction and requires very little computational effort.

3.5 Conclusions

In this chapter we have examined a system in which a single signal is contaminated by several noise sources to generate a number of noisy observations of that signal.

The underlying signal was modelled as an AR process. Based on the assumption that the interfering sources are independent we showed that the noise spectra may be estimated from the noisy observations. A Bayesian method for parameter estimation of the underlying AR process was shown to be feasible in principle, but also highly computationally intensive.

The estimation of the noise spectra was shown to be effective by the demonstration of a broadband noise reduction of several copies of a gramophone disc.

Ensemble-AR (E-AR) Model

4.1	Ensemble-AR Parameter Estimation	57
4.1.1	Covariance Method	57
4.1.2	Correlation Method	58
4.2	Interpolation of Missing Data	59
4.2.1	Single-Channel Interpolation	59
4.2.2	Enhanced Interpolator for Two-Channel Systems . . .	61
4.2.3	Interpolator for Multi-Channel Systems	62
4.2.4	Verification of Interpolation Algorithms	63
4.3	Impulsive Noise Detection	65
4.3.1	Introduction	65
4.3.2	Single-Channel Probabilistic Detector	67
4.3.3	A-Posteriori Detector	68
4.3.4	Bayes' Risk	69
4.4	Two-Channel Impulsive Noise Detection	72
4.4.1	Two-Channel A-Posteriori Detector	72
4.4.2	Analysis of Two-Channel Detector	73
4.4.3	Computational Considerations	76
4.4.4	Two-Channel Detector Performance	77
4.4.5	Multi-Channel Detector	77

4.4.6	Bayes' Risk	78
4.5	Application to Gramophone Record Restoration	78
4.5.1	Two-channel Replay of a Gramophone Record	78
4.5.2	Model Parameter Estimation	79
4.5.3	Two-Channel Impulsive Noise Detection	79
4.5.4	Two-Channel Interpolation	81
4.5.5	Audio Demonstration	82
4.6	Conclusion	83

4

Ensemble-AR (E-AR) Model

CONSIDER a multi-input, multi-output system in which an all-pole filter is driven by an ensemble of white excitation sources to give the multiple observations, as shown in figure 4.1. The filter parameters are common, so the system constrains all its output signals to have the same power spectrum shape. The excitations, however, are unique, thus allowing the output signals to have different time origins, different amplitudes, and different phase relationships between the various signal components.

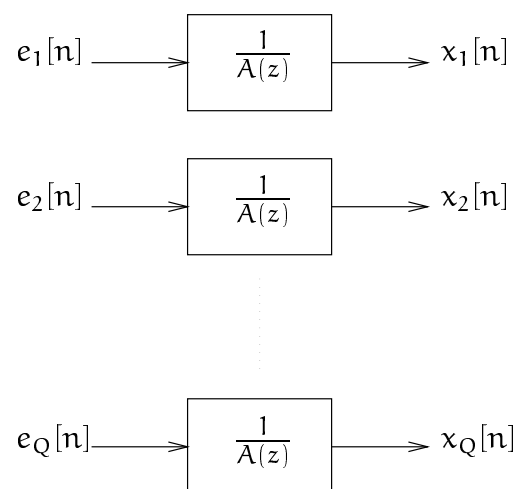


FIGURE 4.1: *Ensemble-AR Signal Model*

For Q signals we may write

$$\mathbf{x}_q(\mathbf{n}) = \mathbf{e}_q(\mathbf{n}) + \sum_{p=1}^P \mathbf{a}(p)\mathbf{x}_q(\mathbf{n} - p), \quad \mathbf{q} = 1 \dots Q, \quad (4.1)$$

or in equivalent matrix notation

$$\underline{\mathbf{x}}_q = \underline{\mathbf{e}}_q + \mathbf{X}_q \underline{\mathbf{a}}, \quad (4.2)$$

where

$$\underline{\mathbf{x}}_q = [\mathbf{x}_q(1) \dots \mathbf{x}_q(i) \dots \mathbf{x}_q(N)]^T, \quad (4.3)$$

$$\underline{\mathbf{e}}_q = [\mathbf{e}_q(1) \dots \mathbf{e}_q(i) \dots \mathbf{e}_q(N)]^T, \quad (4.4)$$

$$\underline{\mathbf{a}} = [\mathbf{a}(1) \dots \mathbf{a}(P)]^T, \quad (4.5)$$

and the i^{th} row of \mathbf{X}_q is $[\mathbf{x}_q(i-1), \dots, \mathbf{x}_q(i-P)]$.

4.1 Ensemble-AR Parameter Estimation

The parameters $\underline{\mathbf{a}}$ may be estimated by a number of means. Two alternatives are described here, which are extensions of the covariance and correlation methods described in section 2.4.

4.1.1 Covariance Method

The covariance method for estimating AR model parameters is outlined in section 2.4.1. In the multi-channel extension we minimise the total excitation energy

$$\mathcal{E} = \sum_{q=1}^Q \underline{\mathbf{e}}_q^T \underline{\mathbf{e}}_q, \quad (4.6)$$

over a finite block of data. Substituting for \mathbf{e}_q from 4.2 we obtain

$$\mathcal{E} = \sum_{q=1}^Q (\underline{\mathbf{x}}_q^T - \underline{\mathbf{a}}^T \mathbf{X}_q^T) (\underline{\mathbf{x}}_q - \mathbf{X}_q \underline{\mathbf{a}}) \quad (4.7)$$

Minimising \mathcal{E} with respect to $\underline{\mathbf{a}}$ by differentiation yields

$$\hat{\underline{\mathbf{a}}} = \left(\sum_{q=1}^Q \mathbf{X}_q^T \mathbf{X}_q \right)^{-1} \left(\sum_{q=1}^Q \mathbf{X}_q^T \underline{\mathbf{x}}_q \right). \quad (4.8)$$

This estimation procedure treats the signals as independent samples from the ensemble of processes AR(P). If the observations \underline{x}_q are truly samples from the ensemble of Gaussian processes AR(P), then so will the excitations \underline{e}_q be white samples from the process $N(0, \sigma_{e_q}^2)$. In the single-channel case where $Q = 1$, equation 4.8 gives the same parameter estimates as the covariance method described in section 2.4.1.

The parameter estimation is robust to power differences between the channels, although altering the amplitude of just some of the signals will change the parameter estimates to some degree. This may be exploited by adjusting the signal amplitudes (in accordance with some *a-priori* knowledge) such that the channels in which we have most trust make a more significant contribution to equation 4.8 than those we distrust.

4.1.2 Correlation Method

For the correlation method we proceed along the same path as far as equation 4.7, but then take the expectation to obtain

$$E[\mathcal{E}] = E \left[\sum_{q=1}^Q (\underline{x}_q^T - \underline{a}^T X_q^T) (\underline{x}_q - X_q \underline{a}) \right] \quad (4.9)$$

and for a block of N samples of each of Q channels

$$\begin{aligned} E[\mathcal{E}] &= E \left[\sum_{q=1}^Q (\underline{x}_q^T - \underline{a}^T X_q^T) (\underline{x}_q - X_q \underline{a}) \right] \quad (4.10) \\ &= (N - P) \sum_{q=1}^Q \sigma_{x_q}^2 + \sum_{q=1}^Q (2 E[\underline{a}^T X_q^T X_q \underline{a}] - E[\underline{x}_q^T X_q \underline{a}] - E[\underline{a}^T X_q^T \underline{x}_q]) \quad (4.11) \end{aligned}$$

Let us define R_q and \underline{r}_q

$$R_q = \begin{bmatrix} r_{x_q x_q}(0) & r_{x_q x_q}(1) & \cdots & r_{x_q x_q}(P-1) \\ r_{x_q x_q}(1) & r_{x_q x_q}(0) & \cdots & r_{x_q x_q}(P-2) \\ \vdots & \vdots & & \vdots \\ r_{x_q x_q}(P-1) & r_{x_q x_q}(P-2) & \cdots & r_{x_q x_q}(0) \end{bmatrix} \quad (4.12)$$

$$\underline{r}_q = \begin{bmatrix} r_{x_q x_q}(1) \\ \vdots \\ r_{x_q x_q}(P) \end{bmatrix} \quad (4.13)$$

$$r_{x_q x_q}(i) = E[x_q[n] x_q[n-i]] \quad (4.14)$$

such that \mathbf{R}_q is the autocorrelation matrix of signal \underline{x}_q , and \underline{r}_q similarly contains auto-correlations of signal \underline{x}_q .

Equation 4.11 may now be rewritten in terms of \mathbf{R}_q and \underline{r}_q

$$\mathcal{E} = (\mathbf{N} - \mathbf{P}) \sum_{q=1}^Q \sigma_{x_q}^2 + \mathbf{P} \sum_{q=1}^Q (2 \underline{\mathbf{a}}^T \mathbf{R}_q \underline{\mathbf{a}} - 2 \underline{\mathbf{r}}_q^T \underline{\mathbf{a}}) \quad (4.15)$$

Minimising \mathcal{E} by differentiation yields the result

$$\hat{\underline{\mathbf{a}}} = \left(\sum_{q=1}^Q \mathbf{R}_q \right)^{-1} \sum_{q=1}^Q \underline{\mathbf{r}}_q \quad (4.16)$$

This system, as for the single-channel case, is Toeplitz and so may be solved efficiently using Levinson-Durbin recursion [61]. Once again the solution for the single-channel case where $Q = 1$ coincides with the standard result for the AR model given in section 2.4.2.

4.2 Interpolation of Missing Data

Suppose that some of the data from one of the channels is missing. If the model parameters are known (or can be reliably estimated from the known data) then this missing data can be interpolated, using the model structure to constrain the nature of the interpolated section.

4.2.1 Single-Channel Interpolation

This single-channel interpolation is due to Vaseghi [101]. It uses the known portion of the data from the corrupted channel and the model parameters to calculate an interpolant which is continuous with the known data either side of the missing data burst.

Suppose we have a single-channel system x_1 , with model parameters $\underline{\mathbf{a}}$, in which there is a burst of L missing data samples. We consider a block of data which comprises these L unknown samples and \mathbf{P} known samples both before and after the corrupted section. The data block under consideration is, therefore, of total length $\mathbf{N} = L + 2\mathbf{P}$ samples.

The excitation sequence for this block may be written as

$$\underline{\mathbf{e}}_1 = \mathbf{A}\underline{\mathbf{x}}_1 \quad (4.17)$$

where

$$\underline{e}_1 = [e_1[P+1] \dots e_1[N]]^T \quad (4.18)$$

$$\underline{x}_1 = [x_1[1] \dots x_1[N]]^T \quad (4.19)$$

and A is the $(N-P)$ by N matrix

$$A = \begin{bmatrix} -a_p & -a_{p-1} & \dots & -a_1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -a_p & -a_{p-1} & \dots & -a_1 & 1 & 0 & \dots & 0 \\ \vdots & & \ddots & \ddots & & \ddots & \ddots & & \vdots \\ 0 & \dots & 0 & -a_p & -a_{p-1} & \dots & -a_1 & 1 & 0 \\ 0 & 0 & \dots & 0 & -a_p & -a_{p-1} & \dots & -a_1 & 1 \end{bmatrix} \quad (4.20)$$

Equation 4.17 may be partitioned so as to separate the samples of \underline{x}_1 into the known samples prior to the corruption (subscript ka), the known samples after the corruption (subscript kb), and the unknown samples (subscript u).

$$\underline{e}_1 = \begin{bmatrix} A_{ka} & A_u & A_{kb} \end{bmatrix} \begin{bmatrix} \underline{x}_{1,ka} \\ \underline{x}_{1,u} \\ \underline{x}_{1,kb} \end{bmatrix} \quad (4.21)$$

Grouping the known samples of \underline{x}_1 together, and permuting the columns of A equivalently gives

$$\underline{e}_1 = \begin{bmatrix} A_{ka} & A_{kb} & A_u \end{bmatrix} \begin{bmatrix} \underline{x}_{1,ka} \\ \underline{x}_{1,kb} \\ \underline{x}_{1,u} \end{bmatrix} \quad (4.22)$$

$$= \begin{bmatrix} A_k & A_u \end{bmatrix} \begin{bmatrix} \underline{x}_{1,k} \\ \underline{x}_{1,u} \end{bmatrix} \quad (4.23)$$

$$= A_k \underline{x}_{1,k} + A_u \underline{x}_{1,u}. \quad (4.24)$$

Vaseghi goes on to minimise $\underline{e}_1^T \underline{e}_1$ with respect to the unknown data samples $\underline{x}_{1,u}$ to give

$$\hat{\underline{x}}_{1,u} = - (A_u^T A_u)^{-1} A_u^T A_k \underline{x}_{1,k} \quad (4.25)$$

provided that the inverse $(A_u^T A_u)^{-1}$ exists.

The interpolant $\hat{\underline{x}}_{1,u}$ is shown to be continuous with the original signal at both ends of the gap, and to display many of the characteristics of the surrounding data.

A significant limitation of this method is that the amplitude of the interpolant tends to decay towards the middle of long gaps [83]. This is a result of the minimisation doing “too good” a job of minimising the error energy. The resulting signal is highly probable, but is not *typical*. By analogy, the most probable observation of a Gaussian variable is its mean, but the ensemble of observations will not typically all be equal to the mean.

4.2.2 Enhanced Interpolator for Two-Channel Systems

In systems where we have two channels (for example, if a monophonic gramophone record is replayed with a stereo pickup—this example is discussed in detail in section 4.5) then it would seem reasonable to use information from a second channel to enhance the performance of the interpolator.

In these cases it is not usually acceptable simply to substitute signal samples from the good channel into the bad one. There are frequently dc-level offsets, low frequency interference, broadband noise and the like which would lead to signal discontinuities if this were attempted. Instead we calculate the excitation signal for the uncorrupted channel, and use this as an estimate for the excitation in the channel we wish to restore.

If we assume that at some time signal \mathbf{x}_2 is uncorrupted, but signal \mathbf{x}_1 contains a burst of L missing samples. In this case we may use the channel 2 excitation

$$\underline{\mathbf{e}}_2 = \mathbf{A}\mathbf{x}_2 \quad (4.26)$$

as an estimate of the true excitation for channel 1, by simply setting $\hat{\underline{\mathbf{e}}}_1 = \underline{\mathbf{e}}_2$.

In this case we may rewrite equation 4.24, subtracting this estimate $\hat{\underline{\mathbf{e}}}_1$ from each side

$$\underline{\mathbf{e}}_1 - \hat{\underline{\mathbf{e}}}_1 = \mathbf{A}_k \underline{\mathbf{x}}_{1,k} + \mathbf{A}_u \underline{\mathbf{x}}_{1,u} - \hat{\underline{\mathbf{e}}}_1 \quad (4.27)$$

We then minimise $(\underline{\mathbf{e}}_1 - \hat{\underline{\mathbf{e}}}_1)^T (\underline{\mathbf{e}}_1 - \hat{\underline{\mathbf{e}}}_1)$ with respect to the unknown samples, obtaining

$$\hat{\underline{\mathbf{x}}}_{1,u} = -(\mathbf{A}_u^T \mathbf{A}_u)^{-1} \mathbf{A}_u^T (\mathbf{A}_k \underline{\mathbf{x}}_{1,k} - \hat{\underline{\mathbf{e}}}_1) \quad (4.28)$$

as the estimate of the missing data.

The inclusion of the excitation estimate transfers information from the good channel to the interpolant, while retaining the benefits of the single-channel interpolator. In particular, the interpolant is guaranteed to be continuous with the known

data either side of the gap. It is, therefore, robust to low frequency interference, dc-level shifts and the like which may differ across the ensemble.

4.2.3 Interpolator for Multi-Channel Systems

A further extension allows inclusion of many excitation estimates in optimal proportion. Suppose that in a Q -channel system there is a burst of missing data in channel 1, but that the corresponding samples in channels 2 to Q are not corrupted. We wish to estimate the data missing from channel 1.

In this case we take a weighted sum of the excitations for all of the uncorrupted channels,

$$\hat{\underline{e}} = \sum_{q=2}^Q \alpha_q \underline{e}_q \quad (4.29)$$

$$= \begin{bmatrix} \underline{e}_2 & \cdots & \underline{e}_Q \end{bmatrix} \begin{bmatrix} \alpha_2 \\ \vdots \\ \alpha_Q \end{bmatrix} \quad (4.30)$$

$$= \underline{E} \underline{\alpha} \quad (4.31)$$

and use this as the excitation estimate for the unknown channel.

The excitation for channel 1 is split into known and unknown parts, as before, and expressed

$$\underline{e}_1 = \underline{A}_k \underline{x}_{k,1} + \underline{A}_u \underline{x}_{u,1} \quad (4.32)$$

and as for the two-channel case the excitation estimate is subtracted from each side to give

$$\underline{e}_1 - \hat{\underline{e}} = \underline{A}_k \underline{x}_{k,1} + \underline{A}_u \underline{x}_{u,1} - \hat{\underline{e}} \quad (4.33)$$

$$\underline{e}_1 - \hat{\underline{e}} = \underline{A}_k \underline{x}_{k,1} + \underline{A}_u \underline{x}_{u,1} - \underline{E} \underline{\alpha}. \quad (4.34)$$

We now group the unknown signal samples $\underline{x}_{u,1}$ with the unknown weights $\underline{\alpha}$ into a single column vector to obtain

$$\underline{e}_1 - \hat{\underline{e}} = \underline{A}_k \underline{x}_{k,1} + \begin{bmatrix} \underline{A}_u & -\underline{E} \end{bmatrix} \begin{bmatrix} \underline{x}_{u,1} \\ \underline{\alpha} \end{bmatrix} \quad (4.35)$$

$$\underline{e}_1 - \hat{\underline{e}} = \underline{A}_k \underline{x}_{k,1} + \underline{M} \begin{bmatrix} \underline{x}_{u,1} \\ \underline{\alpha} \end{bmatrix}. \quad (4.36)$$

Finally we minimise $(\underline{e}_1 - \hat{\underline{e}})^T(\underline{e}_1 - \hat{\underline{e}})$ with respect to the vector of unknowns, giving

$$\begin{bmatrix} \hat{\underline{x}}_{u,1} \\ \hat{\underline{\alpha}} \end{bmatrix} = -(\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \mathbf{A}_k \underline{\mathbf{y}}_{k,1} \quad (4.37)$$

This interpolant possesses all the properties of the two-channel interpolation of section 4.2.2, and adds two significant benefits:

- it allows incorporation of an excitation estimate derived from more than one alternative channel, and
- it is robust to a scale-factor difference between the original signals.

This method can be used directly in place of the two-channel algorithm described above, where it adds the benefit of scale-factor robustness.

4.2.4 Verification of Interpolation Algorithms

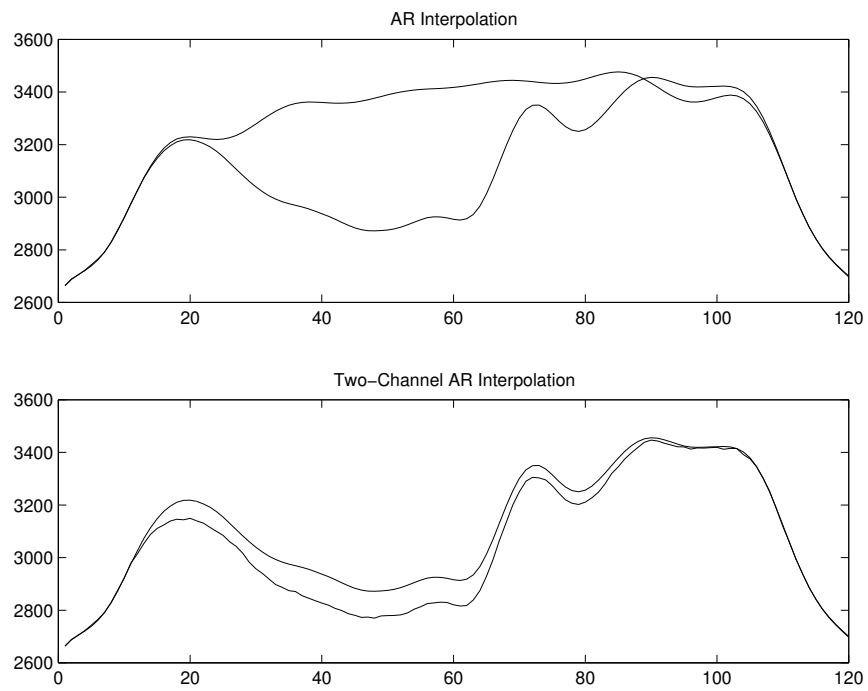
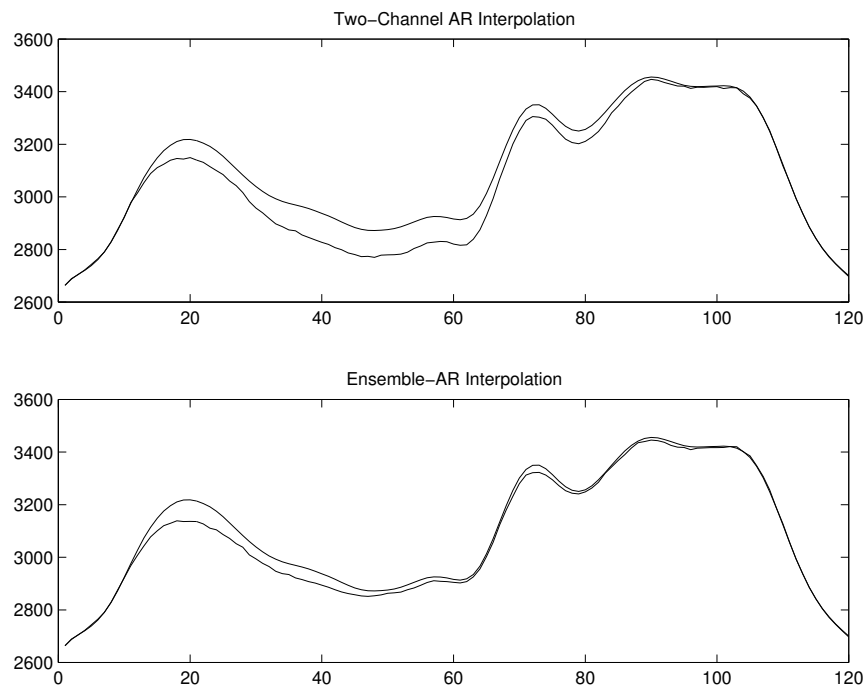
Four channels of synthetic data were generated from an AR model, and each was contaminated with an independent interference signal. One hundred samples from channel 1 were treated as missing, and interpolated using each of the following schemes: single-channel AR (section 4.2.1), two-channel AR (section 4.2.2), ensemble-AR (section 4.2.3).

Figure 4.2 shows the comparison between a single-channel interpolation and a two-channel interpolation that uses an excitation estimate taken directly from one of the other channels. The two-channel interpolation is clearly superior, retaining much more of the character of the original signal than the single-channel interpolator.

Figure 4.3 shows the superior result from using all three other channels to provide the excitation estimate, using the system described in section 4.2.2.

Finally, the robustness of the Ensemble-AR interpolator, compared with the two-channel interpolator, is demonstrated in figure 4.4. The true excitation for the corrupt channel was halved, and then used as the estimate for the two-channel method

$$\hat{\underline{e}}_1 = \frac{1}{2} \underline{e}_1 \quad (4.38)$$

FIGURE 4.2: *Two-Channel Interpolation*FIGURE 4.3: *Ensemble-AR Interpolation*

and this same estimate was used in the Ensemble-AR interpolator. The optimal weight $\alpha = 2.0$ was correctly determined automatically by the Ensemble-AR interpolator.

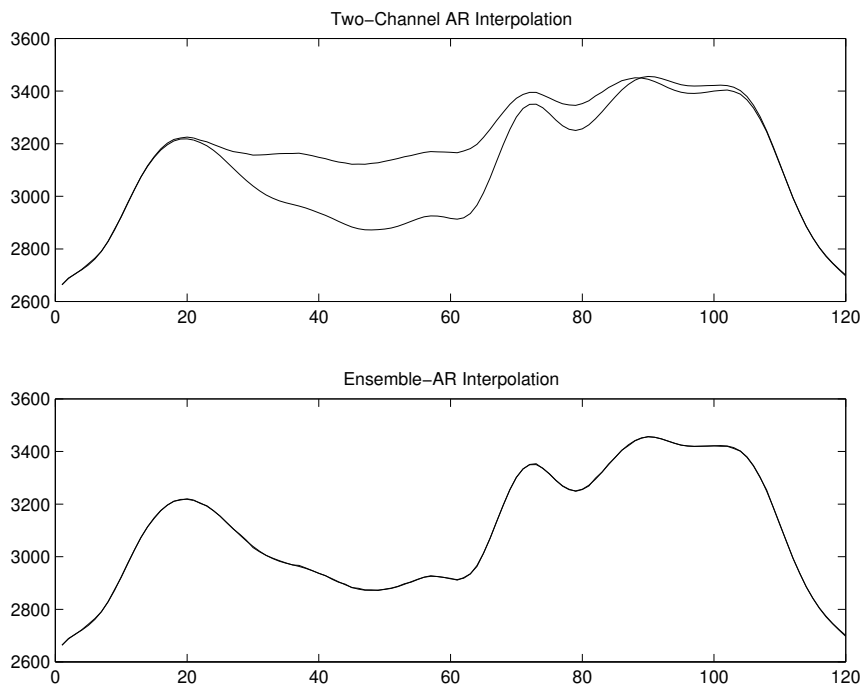


FIGURE 4.4: *Interpolations with non-optimal excitation estimate amplitude*

4.3 Impulsive Noise Detection

The AR model has been used for some time as the basis for detection of impulsive noise in musical signals [101, 35, 83]. This section outlines a basic single-channel probabilistic impulsive noise detector of this type. Section 4.4 describes a new extension of this detector to two-channel and multi-channel Ensemble-AR systems.

4.3.1 Introduction

We model the impulsive interference as zero-mean substitutive Gaussian noise. Thus there are two random processes to consider; that which generates the noise itself, and that which determines whether a given observed signal sample is a true signal sample or whether it is an interference sample.

This model is shown diagrammatically in figure 4.5; the noise signal $N(0, \sigma_n^2)$ is represented by $N(z)$, and the true signal is given by the AR process $X(z) = E_t(z)/A(z)$. The switching is accomplished by the random binary signal $S(z)$, and this results in the observed signal $Y(z)$. We may filter $Y(z)$ with the inverse AR filter $A(z)$ to give an observed excitation sequence $E(z)$.

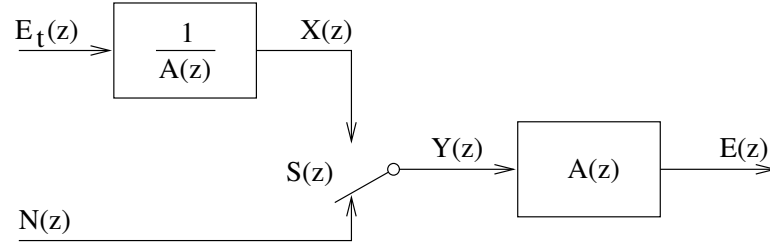


FIGURE 4.5: *Model for Signal corrupted by Substitutive Impulsive Noise*

Let us assume that we have access to the true signal model $A(z)$. In practice we will have an estimate of this model based on known good data, or data that is known to be only mildly contaminated with impulsive noise.

The calculation of $E(z)$ can be implemented as an FIR filtering operation with the model coefficients

$$e[n] = - \sum_{i=0}^P a_i x[n-i] \quad (4.39)$$

where $a_0 = -1$.

Since the model parameters a_i are optimal to whiten the true signal we can assume that the filtered noise $N(z)A(z)$ is large compared with the true excitation $E_t(z)$. Under this assumption we may model a filtered error burst as Gaussian substitutive noise $N(0, \sigma_d^2)$ where $\sigma_d^2 \approx \sigma_n^2$ since $|a_0| = 1$. The correlation of the noise is altered by the filtering operation, but this need not presently concern us since the simple detectors presented here make no assumptions about this correlation.

For impulsive noise in musical signals some 30–40 dB of increased noise/signal separation is achieved by this filtering [101].

The detection process may now be defined as determining whether a given excitation sample is drawn from the true excitation distribution $p_1(e) = N(0, \sigma_e^2)$, or from the interference distribution $p_2(e) = N(0, \sigma_d^2)$. These distributions are plotted in figure G.1.

There is a “smearing” of impulses in the observed excitation which has been noted by a number of authors ([101, 35]). This results in practice in the end of each noise burst being less accurately determined than the start, but this problem is readily overcome by consideration of both the forward and backward observed excitation sequences.

4.3.2 Single-Channel Probabilistic Detector

Suppose that some proportion p_c of the total number of samples is drawn from p_2 . Assume also that they are randomly and uniformly distributed, and that this proportion p_c is unknown. A threshold e_t may be chosen, and the samples $e[n]$ classified as follows:

- if $|e[n]| > e_t$ then sample n is flagged as corrupt, or
- if $|e[n]| < e_t$ then sample n is flagged as valid.

This process is repeated independently for each of the channels of observed data.

The threshold e_t is chosen to minimise the probability of misclassification. In the absence of further *a-priori* information it is given by setting the p.d.f's to be equal and solving for e .

$$e_t^2 = \frac{\sigma_e^2 \sigma_d^2}{\sigma_d^2 - \sigma_e^2} \ln \left(\frac{\sigma_d^2}{\sigma_e^2} \right) \quad (4.40)$$

The misclassification probabilities depend upon the ratio of the variances, and are given by

$$p_m(e_t) = 2 \int_0^{e_t} p_2(e) de \quad (4.41)$$

$$p_f(e_t) = 2 \int_{e_t}^{\infty} p_1(e) de \quad (4.42)$$

where p_m and p_f are the probabilities of a “miss” and a “false-alarm” respectively, defined as follows:

- $p_m(T)$ is the probability that a sample $e[n]$ drawn from distribution p_2 , when thresholded at threshold T , is incorrectly classified as being drawn from p_1
- $p_f(T)$ is the probability that a sample $e[n]$ drawn from distribution p_1 , when thresholded at threshold T , is incorrectly classified as being drawn from p_2 .

The total probability of misclassification of a given sample is given by

$$p_t(\mathbb{T}, p_c) = (1 - p_c) p_f(\mathbb{T}) + p_c p_m(\mathbb{T}) \quad (4.43)$$

where p_c is the proportion of samples that are genuinely drawn from p_2 . These probabilities are evaluated for various ratios σ_d^2/σ_e^2 in table 4.1.

$\sigma_d^2, \sigma_e^2 = 1$	e_t	$p_f(e_t)$	$p_m(e_t)$	$p_t(e_t, p_c), p_c = 5\%$
2	1.177	23.9%	59.5%	25.7%
10	1.600	11.0%	38.7%	12.4%
20	1.776	7.58%	30.9%	8.74%
100	2.157	3.10%	17.1%	3.08%
200	2.308	2.10%	13.0%	2.65%
1000	2.630	0.855%	6.63%	1.14%
2000	2.758	0.582%	4.91%	0.799%
10000	3.035	0.241%	2.42%	0.350%

TABLE 4.1: Misclassification Probabilities for Basic Detector

4.3.3 A-Posteriori Detector

The performance of this detector can be improved by incorporating some additional statistics about the substitutive noise process. This information may be included as priors, and a detector derived which minimises the *a-posteriori* probability of misclassification.

Suppose that we know or can estimate by some independent means the proportion $0 < p_c < 1$ of samples $e[n]$ that are corrupted and therefore drawn from p_2 . We can then say that the *a-priori* probability $\text{pr}(p_2)$ that a given sample is drawn from p_2 is

$$\text{pr}(p_2) = p_c. \quad (4.44)$$

Once we have made the measurement of a given sample value $e[n]$ Bayes' Rule

states that $\text{pr}(p_2 | e[n])$ is given by

$$\text{pr}(p_2 | e[n]) = \frac{\text{pr}(e[n] | p_2) \text{pr}(p_2)}{\text{pr}(e[n])} \quad (4.45)$$

$$= \frac{p_2(e[n]) \text{pr}(p_2)}{p_1(e[n])\text{pr}(p_1) + p_2(e[n])\text{pr}(p_2)} \quad (4.46)$$

$$= \frac{p_2(e[n]) p_c}{p_1(e[n])(1 - p_c) + p_2(e[n])p_c} \quad (4.47)$$

which is the *a-posteriori* probability that sample $e[n]$ is drawn from distribution p_2 , *i.e.* that it is corrupt. These probabilities ($\sigma_e^2 = 1$, $\sigma_d^2 = 100$, $p_c = 5\%$) are plotted in figure G.1.

The posterior probabilities lead to a new classification threshold

$$e_p^2 = \frac{\sigma_e^2 \sigma_d^2}{\sigma_d^2 - \sigma_e^2} \ln \left(\frac{\sigma_d^2 (1 - p_c)^2}{\sigma_e^2 p_c^2} \right) \quad (4.48)$$

which gives miss and false-alarm rates

$$p_m(e_p) = 2 \int_0^{e_p} p_2(e) de \quad (4.49)$$

$$p_f(e_p) = 2 \int_{e_p}^{\infty} p_1(e) de \quad (4.50)$$

$$p_t(e_p, p_c) = p_c p_m(e_p) + (1 - p_c) p_f(e_p) \quad (4.51)$$

where p_m , p_f and p_t are defined as before. Misclassification rates are tabulated in table 4.2, and it can be seen from the table that these are reduced over the simpler detector of table 4.1.

4.3.4 Bayes' Risk

It is interesting to note that for σ_d^2/σ_e^2 close to unity then the miss rate for the *a-posteriori* detector approaches 100%. This is because for relatively infrequent degradation (in our example just 5% of the samples are from the second distribution) the lowest misclassification rate is achieved by simply assuming that all the samples are from the *a-priori* more probable distribution. In contrast, table 4.1 shows that the basic detector has a much worse overall misclassification rate, but that the probability of missing a corrupt sample is much lower.

In a context where there is a limit on the maximum miss or false alarm rate that is acceptable then further constraints may be included in the derivation of the

detector which will put a limit on the range of values adopted for the threshold. For example we could derive a detector with a maximum 1% miss rate by setting

$$p_m(T) = 2 \int_0^T p_2(e) de < 0.01 \quad (4.52)$$

from the definition of p_m (equation 4.41) and deriving from this a constraint on T .

A more general method for optimising the detector is to define a loss function $\lambda(s_i, s_j)$ with estimating the system state to be s_i when the true state is s_j . This provides a flexible framework within which we can penalise each type of misclassification independently. The loss associated with choosing the correct state is taken as $\lambda(s_j, s_j) = 0$.

We then define the *risk* as the expected loss associated with estimating the state to be s_i given the observed data, given by

$$\rho(s_i | \mathbf{e}[\mathbf{n}]) = \sum_{j=1}^{N_s} \lambda(s_i, s_j) \text{pr}(s_j | \mathbf{e}[\mathbf{n}]) \quad (4.53)$$

where N_s is the number of possible system states.

The *Bayes' Risk* is then defined by

$$\mathcal{R} = \sum_{i=1}^{N_s} \rho(s_i | \mathbf{e}[\mathbf{n}]) \quad (4.54)$$

$$= \sum_{i=1}^{N_s} \sum_{j=1}^{N_s} \lambda(s_i | s_j) \text{pr}(s_j | \mathbf{e}[\mathbf{n}]) \quad (4.55)$$

and the optimal detector is the one which minimises \mathcal{R} .

In the simple detector described above we have just two system states, so the Bayes' Risk function simplifies to

$$\mathcal{R} = \lambda(s_2, s_1) \text{pr}(s_1 | \mathbf{e}[\mathbf{n}]) + \lambda(s_1, s_2) \text{pr}(s_2 | \mathbf{e}[\mathbf{n}]) \quad (4.56)$$

which may be minimised with respect to the decision threshold T .

If $\lambda(s_2, s_1) = \lambda(s_1, s_2)$ then this procedure results in the same threshold as for the *a-posteriori* detector given by equation 4.48. A non-uniform loss function causes the detection threshold to be adjusted one way or the other so as to minimise the Bayes' Risk, as opposed to minimising the misclassification probability.

$\sigma_d^2, \sigma_e^2 = 1$	e_p	$p_f(e_p)$	$p_m(e_p)$	$p_t(e_p, p_c), p_c = 5\%$
2	3.628	0.029%	99.0%	4.98%
10	3.017	0.255%	66.0%	3.54%
20	3.058	0.223%	50.6%	2.74%
100	3.256	0.113%	25.5%	1.38%
200	3.353	0.080%	18.7%	1.01%
1000	3.579	0.035%	9.01%	0.48%
2000	3.674	0.024%	6.54%	0.35%
10000	3.886	0.010%	3.10%	0.16%

TABLE 4.2: *A-Posteriori Misclassification Probabilities for Detector*

4.4 Two-Channel Impulsive Noise Detection

In the preceding sections we have been treating the excitation samples individually and in isolation. If, in a two-channel system, we group them into pairs corresponding to the two channels at the same sampling instant and consider their joint distribution then an impulsive noise detector of superior performance may be derived.

4.4.1 Two-Channel A-Posteriori Detector

Once again we assume that we know *a-priori* the proportion p_c of samples that is corrupt, and we also assume that the time-distribution of the impulsive noise is independent in the two channels.

Let us define the vector sample

$$\underline{e} = \begin{bmatrix} e_1[n] \\ e_2[n] \end{bmatrix} \quad (4.57)$$

where each of $e_1[n]$ and $e_2[n]$ may be drawn from either of distributions p_1 or p_2 . There are then four bi-variate distributions from which \underline{e} may be drawn, corresponding to the cases where:

- neither channel is corrupt; e_1 and e_2 both drawn from p_1
- channel 2 is corrupt; e_1 is drawn from p_1 , and e_2 from p_2
- channel 1 is corrupt; e_1 is drawn from p_2 , and e_2 from p_1
- both channels are corrupt; e_1 and e_2 both drawn from p_2 .

Let us call these bi-variate distributions p_{11} , p_{12} , p_{21} , and p_{22} respectively. The two-channel impulsive noise detection procedure may now be defined as determining from which of these four distributions is drawn each sample \underline{e} .

The covariance matrices C_{11} , C_{12} , C_{21} and C_{22} are

$$C_{11} = \begin{bmatrix} \sigma_e^2 & \alpha\sigma_e^2 \\ \alpha\sigma_e^2 & \sigma_e^2 \end{bmatrix} \quad (4.58)$$

$$C_{12} = \begin{bmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_d^2 \end{bmatrix} \quad (4.59)$$

$$C_{21} = \begin{bmatrix} \sigma_d^2 & 0 \\ 0 & \sigma_e^2 \end{bmatrix} \quad (4.60)$$

$$C_{22} = \begin{bmatrix} \sigma_d^2 & \beta\sigma_d^2 \\ \beta\sigma_d^2 & \sigma_d^2 \end{bmatrix} \quad (4.61)$$

where σ_e^2 and σ_d^2 are the variances of the true excitation and corrupted excitation samples respectively, and α and β reflect the degree of correlation between the channels. The zero terms reflect the fact that the degradation is assumed independent of the true signal.

Based on the initial assumptions and the proportion of corrupt samples p_c we may determine the *a-priori* probabilities for each of the distributions as

$$\text{pr}(p_{11}) = (1 - p_c)^2 \quad (4.62)$$

$$\text{pr}(p_{12}) = (1 - p_c) p_c \quad (4.63)$$

$$\text{pr}(p_{21}) = (1 - p_c) p_c \quad (4.64)$$

$$\text{pr}(p_{22}) = p_c^2. \quad (4.65)$$

Using Bayes' Rule we may derive

$$\text{pr}(p_i | \underline{e}) = \frac{\text{pr}(\underline{e} | p_i) \text{pr}(p_i)}{\text{pr}(\underline{e})} \quad (4.66)$$

$$= \frac{p_i(\underline{e}) \text{pr}(p_i)}{\sum_k p_k(\underline{e}) \text{pr}(p_k)}, \quad k = \{11, 12, 21, 22\} \quad (4.67)$$

which is the *a-posteriori* probability that \underline{e} is drawn from distribution i .

Evaluating equation 4.67 for each of the candidate distributions

$$i = \{11, 12, 21, 22\} \quad (4.68)$$

and choosing the most likely enables the sample \underline{e} to be classified.

4.4.2 Analysis of Two-Channel Detector

For the purposes of illustration the parameters for the distributions used for the figures in this section are (unless stated otherwise) $\sigma_e^2 = 1$, $\sigma_d^2 = 100$, $\alpha = 0.8$,

$\beta = -0.8$. These are reasonable figures for the application described in section 4.5 except that the ratio of σ_d/σ_e has been lowered to improve clarity of the figures. The most significant effect of this is that the real-world performance of the detector will be somewhat better than the figures and tables in this section would suggest. As before the *a-priori* probability p_c is set to 5%.

The posterior probabilities $\text{pr}(p_i | \underline{e})$ given by equation 4.67 are plotted in figure G.2 against the two components of the vector excitation sample \underline{e} . The classification of the sample \underline{e} is made by determining which of these probabilities is largest for that sample value.

Figure G.3 is the combination of the four parts of figure G.2 viewed from directly above, and shows more clearly the decision boundaries for this detector.

The boundaries are seen to have a complex shape, which may be analysed further by considering the values of \underline{e} at which the classification changes from distribution p_i to distribution p_j . This gives a line which splits the e -plane into the regions

- $\mathcal{R}_{p_i > p_j}$ in which distribution p_i is more probable than p_j , and
- $\mathcal{R}_{p_i < p_j}$ in which distribution p_i is less probable.

Equation 4.67 gives the posterior probabilities as

$$\text{pr}(p_i | \underline{e}) = \frac{p_i(\underline{e}) \text{pr}(p_i)}{\sum_k p_k(\underline{e}) \text{pr}(p_k)} \quad (4.69)$$

$$\text{pr}(p_j | \underline{e}) = \frac{p_j(\underline{e}) \text{pr}(p_j)}{\sum_k p_k(\underline{e}) \text{pr}(p_k)} \quad (4.70)$$

for $k = \{11, 12, 21, 22\}$.

Setting these equal gives

$$\frac{p_i(\underline{e}) \text{pr}(p_i)}{\sum_k p_k(\underline{e}) \text{pr}(p_k)} = \frac{p_j(\underline{e}) \text{pr}(p_j)}{\sum_k p_k(\underline{e}) \text{pr}(p_k)} \quad (4.71)$$

$$p_i(\underline{e}) \text{pr}(p_i) = p_j(\underline{e}) \text{pr}(p_j) \quad (4.72)$$

$$\text{pr}(p_i) \frac{1}{2\pi|C_i|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\underline{e}^T C_i^{-1} \underline{e}\right) = \text{pr}(p_j) \frac{1}{2\pi|C_j|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\underline{e}^T C_j^{-1} \underline{e}\right) \quad (4.73)$$

$$\frac{\text{pr}(p_i) |C_j|^{\frac{1}{2}}}{\text{pr}(p_j) |C_i|^{\frac{1}{2}}} = \exp\left(-\frac{1}{2}\underline{e}^T (C_i^{-1} - C_j^{-1}) \underline{e}\right) \quad (4.74)$$

Squaring both sides of equation 4.74 (noting that it is guaranteed by its form to

be positive) and then taking its logarithm reduces it to a quadratic in \underline{e} .

$$\ln \left(\frac{\text{pr}(\mathbf{p}_i)^2 |C_j|}{\text{pr}(\mathbf{p}_j)^2 |C_i|} \right) = -\underline{e}^\top (C_i^{-1} - C_j^{-1}) \underline{e} \quad (4.75)$$

$$\ln \left(\frac{\text{pr}(\mathbf{p}_i)^2 |C_j|}{\text{pr}(\mathbf{p}_j)^2 |C_i|} \right) = \underline{e}^\top (C_j^{-1} - C_i^{-1}) \underline{e} \quad (4.76)$$

The left hand side of equation 4.76 is a scalar k and we also substitute

$$M^{-1} = C_j^{-1} - C_i^{-1} \quad (4.77)$$

to give

$$k = \underline{e}^\top M^{-1} \underline{e} \quad (4.78)$$

which defines an ellipse or hyperbola in the \underline{e} -plane.

Since C_i and C_j are symmetric it follows that M is also symmetric and therefore has an eigenvalue decomposition

$$M = R\Lambda R^\top \quad (4.79)$$

$$M^{-1} = R\Lambda^{-1}R^\top \quad (4.80)$$

where Λ is a diagonal matrix of the eigenvalues of M , and R is an orthogonal matrix consisting of the corresponding eigenvectors.

Making this substitution, equation 4.78 becomes

$$k = \underline{e}^\top R\Lambda^{-1}R^\top \underline{e} \quad (4.81)$$

$$k = \underline{u}^\top \Lambda^{-1} \underline{u} \quad (4.82)$$

whereupon it can be seen that R^\top represents the rotation from \mathbf{e} -space to \mathbf{u} -space, and Λ^{-1} is a simple scaling matrix.

The region of the \mathbf{e} -plane over which the sample is classified to be from distribution \mathbf{p}_i is the intersection of three such ellipses/hyperbolae, each being the boundary with one of the other candidate distributions. For example, the sample is classified as being from distribution \mathbf{p}_{11} , *i.e.* uncorrupted in both channels, in the region

$$\mathcal{R}_{\mathbf{p}_{11}} = \mathcal{R}_{\mathbf{p}_{11} > \mathbf{p}_{22}} \cap \mathcal{R}_{\mathbf{p}_{11} > \mathbf{p}_{21}} \cap \mathcal{R}_{\mathbf{p}_{11} > \mathbf{p}_{12}}. \quad (4.83)$$

This region is illustrated in figure G.4, and is seen to match the shape of the central region in figure G.3.

4.4.3 Computational Considerations

The one-channel threshold detector is very cheap, whereas the two-channel *a-posteriori* detector requires the evaluation of 4.67 (which includes an $\exp(\cdot)$ function and a division) four times to calculate each of the posterior probabilities.

Possible optimisations include:

- use of log probabilities instead of true probabilities,
- devising of a simple derived test that passes for the vast majority of samples that are non-corrupt in both channels,

and we examine each of these in turn.

4.4.3.1 Log Probabilities

Taking the logarithm of equation 4.67 we obtain

$$\ln(\text{pr}(\mathbf{p}_i | \underline{\mathbf{e}})) = \ln(\mathbf{p}_i(\underline{\mathbf{e}})) + \ln(\text{pr}(\mathbf{p}_i)) - \ln\left(\sum_k \mathbf{p}_k(\underline{\mathbf{e}})\text{pr}(\mathbf{p}_k)\right) \quad (4.84)$$

$$= -\frac{1}{2}\underline{\mathbf{e}}^T \mathbf{C}_i \underline{\mathbf{e}} + l_i - \ln\left(\sum_k \mathbf{p}_k(\underline{\mathbf{e}})\text{pr}(\mathbf{p}_k)\right) + k \quad (4.85)$$

$$= -\frac{1}{2}\underline{\mathbf{e}}^T \mathbf{C}_i \underline{\mathbf{e}} + l_i - L(\underline{\mathbf{e}}) + k \quad (4.86)$$

where $l_i = \ln(\text{pr}(\mathbf{p}_i))$ is the log prior probability of distribution i which may be pre-calculated once. The log-evidence $L(\underline{\mathbf{e}})$ and the scalar constant k may be calculated once, and then equation 4.86 is evaluated for each candidate distribution. The requirement to calculate four exponential functions per excitation sample has been eliminated. Furthermore, if just a detector is required (as opposed to a complete evaluation of the posterior p.d.f's) then there is no need to calculate either the log evidence $L(\underline{\mathbf{e}})$ or the scaling constant k .

4.4.3.2 Derived Test

Figure G.4 shows the non-trivial region $\mathcal{R}_{\mathbf{p}_{11}}$ in which the majority of the excitation samples fall. From the component equations of this boundary it may be possible to calculate, for example, an ellipse of maximum area which fits wholly within region $\mathcal{R}_{\mathbf{p}_{11}}$. The test for whether a given sample $\underline{\mathbf{e}}$ falls within this ellipse will be relatively straightforward, and this will be satisfied for the majority of

the samples. Only when a sample falls outside the ellipse will all of the posterior probabilities need to be calculated.

The \mathbf{u} -space defined by the rotation $\mathbf{u} = \mathbf{R}^T \mathbf{e}$ may well be useful in the derivation of such a test; the test for a sample to be inside an ellipse whos axes are aligned with the co-ordinate system is relatively straightforward. The details of this derived test are left as an issue which requires further research, and which would become particularly important in a multi-channel system. For a two-channel, non-real-time system it is viable to use the log posterior probabilities given by equation 4.86.

4.4.4 Two-Channel Detector Performance

Because the regions in the \mathbf{e} -plane over which the integrals need to be evaluated are difficult we rely on simulations to assess the performance of this detector. The results of these simulations show that for data fitting the model the two-channel detector gives superior results (table 4.3) to the single-channel detector (table 4.2).

$\sigma_d^2, \sigma_e^2 = 1$	$p_f(\mathbf{e}_p)$	$p_m(\mathbf{e}_p)$	$p_t(\mathbf{e}_p, \mathbf{p}_c), \mathbf{p}_c = 5\%$
2	2.86%	51.1%	4.54%
10	1.48%	32.0%	2.68%
20	1.05%	25.0%	2.04%
100	0.418%	13.6%	0.974%
200	0.291%	10.1%	0.711%
1000	0.123%	4.95%	0.338%
2000	0.072%	3.56%	0.239%
10000	0.036%	1.70%	0.112%

TABLE 4.3: *A-Posteriori Misclassification for Two-Channel Detector*

4.4.5 Multi-Channel Detector

Extensions are clearly possible to make a multi-channel detector. With Q channels there are Q^2 distributions to choose from, each having Q variables. The mathematics concerning the decision boundaries is a direct extension of the two-channel system. Because of the increased dimensionality the computational optimisations described in section 4.4.3 will become highly important.

4.4.6 Bayes' Risk

It is clearly possible, and will often be desirable, to use the concept of Bayes' Risk and a non-uniform loss function in a practical application of either the two-channel or multi-channel detector.

4.5 Application to Gramophone Record Restoration

The ensemble model has been used successfully to enhance the algorithms presented by Vaseghi [101] for gramophone record restoration. Two signals are read from a monophonic record and the additional information that this provides over a single-channel transcription allows improvements to be made both in the detection of impulsive noise, and in the subsequent interpolation of the missing data.

This application has been summarised in [48].

4.5.1 Two-channel Replay of a Gramophone Record

A conventional stereo record-player cartridge may be used to sample the two groove walls of a monophonic gramophone record independently, providing two spatial samples of the stored signal. Such a cartridge has two electrical outputs, which ideally correspond to stylus motion along a pair of perpendicular axes at 45° to the vertical, and perpendicular to the longitudinal axis of the cartridge body.

A positive output is produced on the left electrical output by motion along vector L, and on the right output by motion along vector R. The vast majority of 78 rpm records are recorded with a single signal which modulates the horizontal displacement of the groove. It can be seen from figure 4.6 that this motion ideally produces an equal, in-phase output on the electrical outputs corresponding to L and R.

A further possible refinement would be to extract more signals from different depths in the groove by using several styli of varying sizes, or more exotic techniques such as scanning electron microscopy. However, the use of a standard two-channel pickup with a single appropriate stylus represents an inexpensive transcription system and already yields significant advantages over using a single-channel transcription as we shall show.

A few early discs were produced using *hill & dale* modulation, where the signal modulates the groove depth; such recordings produce out-of-phase outputs, but it is a trivial matter to invert one channel to allow for this. Blumlein's patent [10] describes a stereo recording system in which one channel is recorded laterally and the other vertically. Due to implementational difficulties, largely due to the asymmetry of the setup, this format never gained popularity and was dropped for commercial systems in favour of the scheme illustrated above.

4.5.2 Model Parameter Estimation

Figure 4.7 shows a pair of signals recovered from a record by using a stereo pickup. It can be seen that the musical information is very similar in the two channels, but that the impulsive noise is very different.

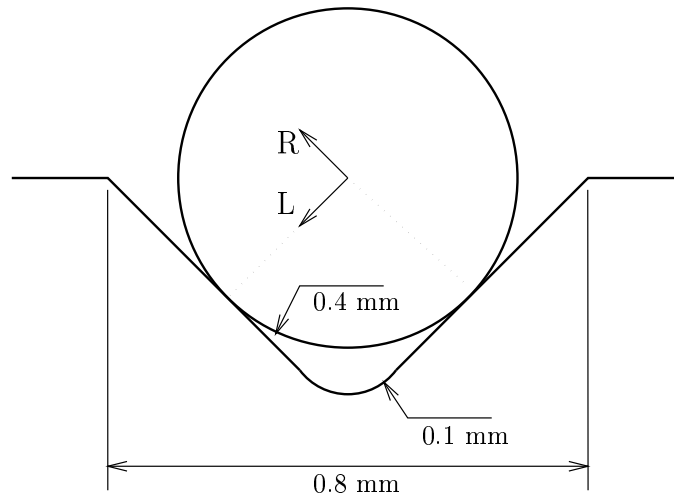
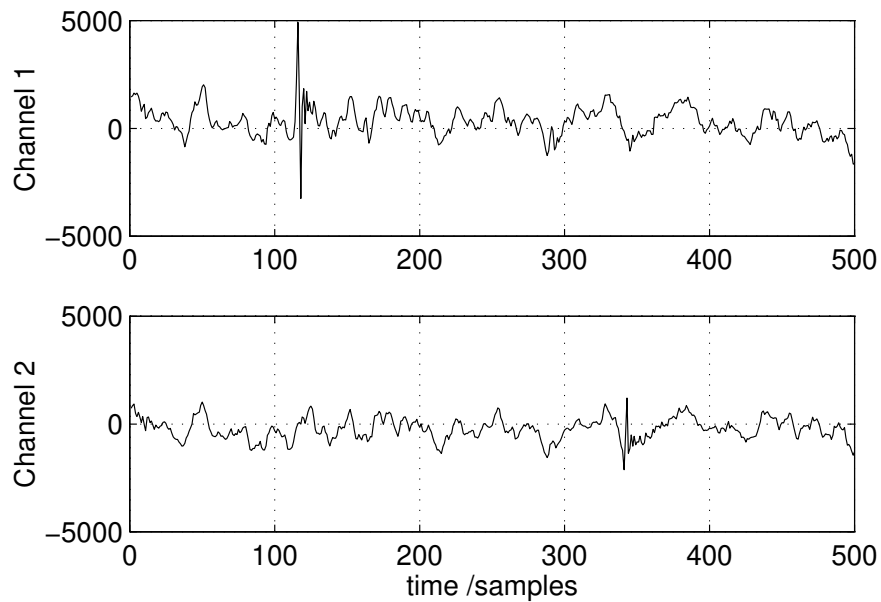
A set of model parameters was estimated from this data using equation 4.8, with the model order P being 25. This figure is chosen somewhat arbitrarily, but proves to be a good compromise between performance and computational complexity for a wide range of audio material.

4.5.3 Two-Channel Impulsive Noise Detection

In order to maximise the performance of the *a-posteriori* two-channel detector in this application we make use of the following observations and assumptions:

- Since both channels represent the same musical data we expect the true excitation signals to be highly correlated.
- For a given type of groove-wall degradation the impulses in one channel will be of opposite polarity to those in the second channel. In other words, if degradation occurs simultaneously on the two channels there will be a negative correlation between the sample values of the two channels.
- The impulsive noise distribution in time is uniform and random, and is independent between the channels.

The first of these observations is almost self-evident because if the two original signals were identical then so would be the two excitation signals. Since we are assuming that the musical information is the same in the two groove walls then we would expect that the part of the excitation signals which represents the music to be equal also. However, noise and errors inherent in the model-estimation process

FIGURE 4.6: *Typical Stylus and Groove Geometry.*FIGURE 4.7: *Signals from a monophonic 78 rpm record*

(largely arising from the presence of impulsive noise) will tend to degrade this equality to some extent, but we assume that we are left with a strong correlation.

The negative correlation of the impulsive-noise samples arises from the fact that, for a given type of simultaneous groove-wall degradation (*e.g.* small amounts of material eroded from the groove walls), the resulting stylus motion is predominantly vertical. With reference to figure 4.6 it can be seen that this results in positive output on one channel, and negative on the other. This polarity inversion is observed in figure 4.7.

The assumption of time-independence between the channels is found to hold adequately true in practice and may also be observed qualitatively in the signals shown in figure 4.7.

4.5.3.1 Practical Considerations

The one-sided nature of the predictor means that impulses in the original signal are smeared in the excitation waveform by the impulse response of the detection filter. In order to determine accurately both the onset and the finish of a particular disturbance it is necessary to apply this filter to both the forward and reverse prediction errors.

4.5.4 Two-Channel Interpolation

For interpolation the data is split into sections which have P uncorrupted samples at each end of both channels; call these \mathbf{y}_1 and \mathbf{y}_2 . The one with the shorter maximum burst length of corrupt samples is interpolated first (if necessary); let us assume that this is channel one.

Any uncorrupted excitation samples from channel two may be used as an estimate of the excitation to be incorporated *via* equation 4.28; the excitation samples that are corrupted in channel two are assumed to be zero. This yields an estimate of the uncorrupted channel one data, $\hat{\mathbf{x}}_1$.

An excitation estimate can be calculated for channel two as

$$\hat{\mathbf{e}}_2 = \mathbf{A}\hat{\mathbf{x}}_1 \quad (4.87)$$

and this can be used with equation 4.28 to give an interpolation of the second, more highly-degraded channel.

Interpolations made in this manner over long gap lengths retain much more of the expected character of the signal than single-channel interpolations. In the

single-channel case the interpolated excitation power tends to drop in the gap as the minimisation does “too good” a job of reducing the excitation energy. This observation stems from the fact that a stream of near-zero excitation samples is highly *likely* as the model excitation, but is not *typical* of the excitation sequences that are encountered in practice.

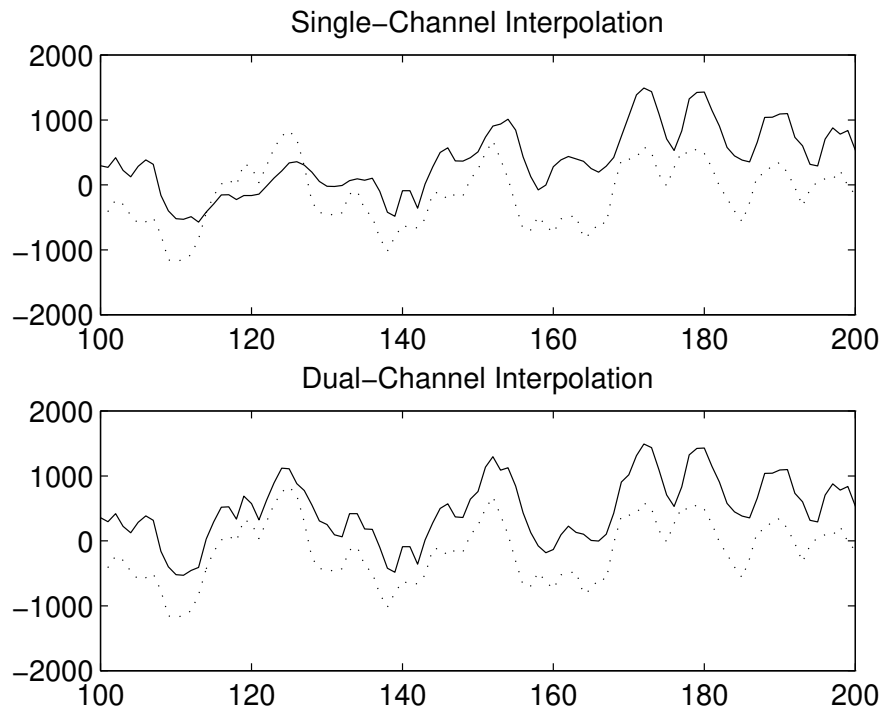


FIGURE 4.8: *Two-Channel Interpolation of Musical Data*

Figure 4.8 shows interpolations of the example two-channel signal of figure 4.7 around the large noise pulse in channel one. The dotted line is the channel two signal, and the solid line is the interpolated estimate of channel one in each case. The two-channel interpolation transfers much of the high-frequency information from channel two into channel one over the interpolated section; this is particularly noticeable around samples 110–135.

4.5.5 Audio Demonstration

Tracks [7] and [8] on the accompanying CD demonstrate the efficacy of the complete two-channel impulsive noise detection and removal algorithm presented in this chapter.

Track [7] is a two-channel transcription of a monophonic disc made in 1935. The impulsive noise is heard to be significantly different in the two channels, this being particularly clear if headphones are used instead of loudspeakers.

This track was restored using the two-channel impulsive noise detector and interpolation algorithms described in this chapter, and the result can be heard on track [8].

4.6 Conclusion

We have extended the AR model to multi-channel systems by treating the signals as members of an ensemble of identical AR processes. We have shown that the parameters of such a model may be estimated from observed data.

The model has been used to detect impulsive outliers in the observed data, and subsequently to interpolate the data samples destroyed by this impulsive interference. The multi-channel model was shown to be superior to the single channel model for both of these operations.

As a demonstration of the practical application of this model and the associated algorithms, the theory was used to develop a restoration algorithm for use on two-channel transcriptions of monophonic gramophone discs. This restoration was demonstrated to be effective on real archive material.

Coupled ARMA (C-ARMA) Model

5.1	Introduction	85
5.2	Application to Stereo Audio Signals	86
5.2.1	Phase Stereo	87
5.2.2	Intensity Stereo	88
5.3	C-ARMA Model Parameter Estimation	88
5.3.1	Estimation of Model Zeros	89
5.3.2	Estimation of Model Poles	91
5.3.3	Verification of Parameter Estimation Algorithm	92
5.3.4	Total Least Squares Zero-Estimation	94
5.4	Interpolation of Missing Data	106
5.4.1	MAP Interpolation of Gaussian Signals	106
5.4.2	Interpolation of Gaussian ARMA Signals	107
5.4.3	Stereo C-ARMA Interpolator	110
5.5	Interpolation Tests	112
5.5.1	Synthetic Data	112
5.5.2	Audio Data	112
5.6	Conclusions	113

Coupled ARMA (C-ARMA) Model

TWO-CHANNEL SIGNALS that are designed to convey the illusion of spatially separated sources when replayed over a pair of loudspeakers or over headphones are referred to as *stereophonic* signals. Such signals form the vast majority of available recorded material, and this is also the most common format for audio data distribution; there is, therefore, a strong motivation for generating good models for applications such as coding, data-reduction and sound analysis and synthesis.

The Coupled-ARMA (C-ARMA) model is put forward here as a possible model for these types of signal, its justification being drawn principally from consideration of the practicalities of typical recording scenarios. It is a general model, applicable to any stereo signal in which there is inter-channel redundancy. We show also how it relates to the special cases of intensity and phase stereo that were described in section 2.7.

5.1 Introduction

The C-ARMA model is shown diagrammatically in figure 5.1. We attempt to model the sound source itself as an AR process, comprising the all-pole filter $\frac{1}{A(z)}$ and a white excitation signal. The propagation transfer functions to the two microphone electrical outputs are modelled as the pair of moving-average filters

$B_L(z)$ and $B_R(z)$.

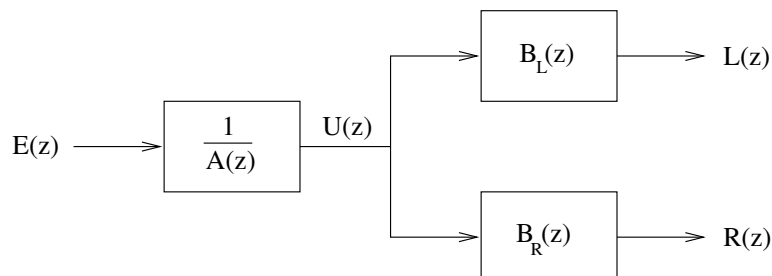


FIGURE 5.1: *General C-ARMA Model for a Stereo Signal*

The system may be described by the equations

$$L(z) = E(z) \frac{1}{A(z)} B_L(z) \quad (5.1)$$

$$R(z) = E(z) \frac{1}{A(z)} B_R(z) \quad (5.2)$$

which describe each of the signal components L and R as an ARMA process, but whose recursive sections (which model the sound source itself) are identical. The excitation is represented by $E(z)$, the z -transform of the excitation samples for a known, finite block of data. For convenience we also define the signal

$$U(z) = E(z) \frac{1}{A(z)} \quad (5.3)$$

which is internal to the model.

5.2 Application to Stereo Audio Signals

The C-ARMA model may be effectively applied to two-channel audio signals when there is significant inter-channel redundancy. Many stereo signals fall into this class since the two channels of information originate in the same acoustic space. Even a signal component that is present in one channel only may be modelled, since poles in the recursive section may be precisely cancelled by zeros in only one of the moving-average sections. This situation is, in fact, frequently encountered, and does not present difficulties for the model parameter estimation problem as is demonstrated later.

The sound source itself is modelled as an autoregressive process as was mentioned previously. It has been shown in previous chapters and extensively in the literature that this model may be successfully applied to audio signals of many types, including music and speech.

Stereo signals rely on both time of arrival differences and amplitude differences between the left and right channels to convey a spatial illusion. Both of these may be applied by the FIR parts of the model. The length of the FIR part needs to be sufficient that this is possible, and in practice an order of 50–80 for this part of the model has been found satisfactory. This allows general shaping of the power spectrum, and allows time of arrival differences of up to 50–80 samples between the channels. This corresponds to up to approximately 1.5ms, or 0.5m path difference for sound in air, which is a realistic figure for acoustically-recorded stereo signals as discussed in section 2.7.

It should be noted that there is a gross time delay associated with the transit time from the sound source to the microphone system, which is not included in the model. This simplification is justified by the fact that the source is assumed to be distant compared with the microphone spacing, and that therefore this delay (from the source to the “acoustic centre” of the microphone system can be assumed to be common to both signals, and hence taken outside of the model for purposes of coding and data interpolation *etc.* If the C-ARMA model were to be employed in a feedback loop (such as, for example in an active noise control system) then this delay is likely to become important and should be included between the recursive and non-recursive model components.

5.2.1 Phase Stereo

Phase stereo records the acoustic pressure at a pair of points which are separated in space, but which are close together compared with the distance to the source. Pressure is a scalar field in 3-D space, and thus the signals from the two microphones have the same power spectrum, but the signals arrive at the microphones at different times according to the microphone spacing and source angle relative to the line joining the microphones.

This scenario is precisely reproducible within the C-ARMA structure. In this special case the FIR filters become all-pass designs and are responsible for reproducing the relative time/phase shifts of the signal components. Since the FIR sections are all-pass, the power spectrum of the left and right signals are both

wholly determined by the AR section of the model. This is common to both signals, so the power spectra at the two outputs are identical.

5.2.2 Intensity Stereo

Intensity stereo records air particle velocity along two axes at one point in space. Particle velocity is a vector field and the recorded signal amplitudes depend on the angle the wavefront makes with the microphone axis. Since in a typical scenario different signals come from many directions the signals recorded may differ in power spectrum, but the times of arrival will be simultaneous for the coincident microphone pair.

This situation is modelled by a C-ARMA system in which the zeros form a pair of minimum-phase filters. These preserve the time of arrival information, which is effectively coded in the excitation signal. The AR section generates a redundant power spectrum, which is then shaped for the individual channels by the separate FIR model sections.

5.3 C-ARMA Model Parameter Estimation

There is no straightforward unimodal solution to the model parameter estimation problem for a general ARMA model, as there is for the AR, MO-AR and E-AR models considered in previous chapters. The C-ARMA may be considered as a pair of ARMA models whose AR sections are identical. From this point of view it may be expected that no straightforward algorithm is available for estimation of the parameters.

In a few special cases, where the recording setup is known and simple, it may be possible to make some estimation of the model parameters $\mathbf{b}_L(i)$ and $\mathbf{b}_R(i)$ from physical principles. However, for the model to be generally useful we need a means of estimating the parameters from the stereo signal itself.

However, the constraint that the recursive sections are identical allows a significant simplification. We may estimate the MA sections first by considering the inter-channel transfer function between the left and right output signals. This allows us to estimate the model zeros for both channels, $B_L(z)$ and $B_R(z)$. Once these filters are known we can generate a pair of signals from which we may estimate $A(z)$ and hence the model poles.

5.3.1 Estimation of Model Zeros

Some manipulation of the z-domain equations 5.1 and 5.2 yields the relationship

$$L(z)B_R(z) = R(z)B_L(z) \quad (5.4)$$

which allows us to express the interchannel function without reference to the originating excitation signal $E(z)$ or the source model $1/A(z)$. Solution of this equation may also be regarded as the determination of $B_R(z)/B_L(z)$ which is the inter-channel transfer function.

If we restrict ourselves to consideration of a causal system whose FIR parts are of order P then we may rewrite 5.4 as the time-domain difference equation

$$\sum_{i=0}^P b_R[i] l[n-i] = \sum_{i=0}^P b_L[i] r[n-i] \quad (5.5)$$

thereby expressing a direct relationship between the samples of one channel and the samples of its partner. In order to make use of this relationship we require knowledge of the model parameters $b_L[i]$ and $b_R[i]$.

With no loss of generality we may set $b_R[0] = 1$. Separating this term from the summation, and introducing a small modelling error ϵ we obtain

$$l[n] + \sum_{i=1}^P b_R[i] l[n-i] = \sum_{i=0}^P b_L[i] r[n-i] + \epsilon[n]. \quad (5.6)$$

Equation 5.6 may be rearranged in two ways. Firstly it may be viewed as a prediction of $l[n]$ based on $r[n]$, and past samples of both channels $l[n-i]$, $r[n-i]$:

$$l[n] = \sum_{i=0}^P b_L[i] r[n-i] - \sum_{i=1}^P b_R[i] l[n-i] + \epsilon[n] \quad (5.7)$$

$$l[n] = \hat{l}[n] + \epsilon[n]. \quad (5.8)$$

In this context $\epsilon[n]$ is regarded as the prediction error, and a successful model will minimise this error.

Rearranging alternatively gives

$$\epsilon[n] = l[n] - \hat{l}[n] \quad (5.9)$$

$$\epsilon[n] = l[n] + \sum_{i=1}^P b_R[i] l[n-i] - \sum_{i=0}^P b_L[i] r[n-i] \quad (5.10)$$

and in this form it can be used to estimate the model parameters $\{b_L, b_R\}$ by minimising the error ϵ over a block of known data.

5.3.1.1 Covariance Method

Expressing 5.10 as a matrix equation for a finite block of stereo samples gives

$$\underline{\epsilon} = \underline{l} + L \underline{b}_R - R \underline{b}_L \quad (5.11)$$

$$\underline{\epsilon} = \underline{l} + \begin{bmatrix} L & -R \end{bmatrix} \begin{bmatrix} \underline{b}_R \\ \underline{b}_L \end{bmatrix} \quad (5.12)$$

$$\underline{\epsilon} = \underline{l} + M \begin{bmatrix} \underline{b}_R \\ \underline{b}_L \end{bmatrix} \quad (5.13)$$

and minimising $\underline{\epsilon}^T \underline{\epsilon}$ yields coefficients $\{\underline{b}_L, \underline{b}_R\}$ given by the solutions to the linear equation

$$(M^T M) \begin{bmatrix} \underline{b}_R \\ \underline{b}_L \end{bmatrix} = -M^T \underline{r}. \quad (5.14)$$

This method is analogous to the covariance method for AR parameter determination that was described in chapter 2. The matrix $M^T M$ contains terms which represent both the auto-covariances and the cross-covariances of the observed signals, as may be seen by splitting equation 5.14 into two sets of equations, one for \underline{b}_L and one for \underline{b}_R . Resubstituting $\begin{bmatrix} L & -R \end{bmatrix} = M$ we obtain

$$\begin{bmatrix} L^T \\ -R^T \end{bmatrix} \begin{bmatrix} L & -R \end{bmatrix} \begin{bmatrix} \underline{b}_R \\ \underline{b}_L \end{bmatrix} = - \begin{bmatrix} L^T \\ -R^T \end{bmatrix} \underline{r} \quad (5.15)$$

$$\begin{bmatrix} L^T L & -L^T R \\ -R^T L & R^T R \end{bmatrix} \begin{bmatrix} \underline{b}_R \\ \underline{b}_L \end{bmatrix} = - \begin{bmatrix} L^T \\ -R^T \end{bmatrix} \underline{r}. \quad (5.16)$$

and solution of this linear system for $\begin{bmatrix} \underline{b}_R^T & \underline{b}_L^T \end{bmatrix}^T$ gives the model parameters.

5.3.1.2 Correlation Method

A similar set of equations analogous to the correlation method may be derived

$$\begin{bmatrix} C_{LL} & -C_{LR} \\ -C_{RL} & C_{RR} \end{bmatrix} \begin{bmatrix} \underline{b}_R \\ \underline{b}_L \end{bmatrix} = - \begin{bmatrix} \underline{c}_{Lr} \\ -\underline{c}_{Rr} \end{bmatrix} \quad (5.17)$$

where C_{xx} and \underline{c}_{xx} are the appropriate correlation matrices and vectors.

This system is not Toeplitz and so Levinson's efficient recursive algorithm [61] is not directly applicable. We can, however, use its block-Toeplitz structure and split equation 5.16 into two parts

$$C_{LL} \underline{b}_R - C_{LR} \underline{b}_L = -\underline{c}_{Lr} \quad (5.18)$$

$$-C_{RL} \underline{b}_R + C_{RR} \underline{b}_L = \underline{c}_{Rr} \quad (5.19)$$

which may be solved simultaneously for the two unknown vectors

$$(C_{RR} - C_{RL} C_{LL}^{-1} C_{LR}) \underline{b}_L = C_{RL} C_{LL}^{-1} \underline{c}_{Lr} + \underline{c}_{Rr} \quad (5.20)$$

$$(C_{LL} - C_{LR} C_{RR}^{-1} C_{RL}) \underline{b}_R = C_{LR} C_{RR}^{-1} \underline{c}_{Rr} + \underline{c}_{Lr}. \quad (5.21)$$

Estimation of the parameters by this correlation method requires the solution of these two Toeplitz systems, and additionally calculation of the inverses C_{LL}^{-1} and C_{RR}^{-1} . For large systems this will be more efficient than direct solution of equation 5.14.

5.3.2 Estimation of Model Poles

To estimate the model poles we wish to invert the zeros of the model given by $B_L(z)$ and $B_R(z)$, and filter the observations $L(z)$, $R(z)$ by these inverses to obtain two estimates of the model internal signal $U(z)$ defined in equation 5.3. From these we can then estimate the model poles by any of the standard AR model parameter estimation techniques.

However, the FIR filters $B_L(z)$ and $B_R(z)$ will not in general be minimum-phase; in fact one of the justifications for the model structure given above is that these filters can model time delays. Therefore the filters do not, in general, have stable inverses and hence direct calculation of $U(z)$ is impossible.

We can transform $B_L(z)$ and $B_R(z)$ into minimum-phase equivalents by reflecting in the unit circle any zeros z_i for which $|z_i| > 1$ by the transformation

$$z'_i = \frac{1}{z_i}. \quad (5.22)$$

This yields a pair of minimum-phase filters $B'_L(z)$, $B'_R(z)$ whose power responses are identical to the power responses of the true model filters:

$$B'_L(e^{j\theta}) B'^{*}_L(e^{j\theta}) = B_L(e^{j\theta}) B^*_L(e^{j\theta}) \quad (5.23)$$

$$B'_R(e^{j\theta}) B'^{*}_R(e^{j\theta}) = B_R(e^{j\theta}) B^*_R(e^{j\theta}). \quad (5.24)$$

The original channel filters $B_L(z)$ and $B_R(z)$ may now be expressed in terms of these minimum phase filters and a pair of all-pass filters which provide the excess phase shift. Thus we may write

$$B_L(z) = B'_L(z) H_L(z) \quad (5.25)$$

$$B_R(z) = B'_R(z) H_R(z) \quad (5.26)$$

where $|H_L(e^{j\theta})| = |H_R(e^{j\theta})| = 1$. These expressions lead to the expanded block diagram shown in figure 5.2.

The minimum-phase filters have stable inverses so we may generate two new signals

$$U_L(z) = L(z) \frac{1}{B'_L(z)} \quad (5.27)$$

$$U_R(z) = R(z) \frac{1}{B'_R(z)} \quad (5.28)$$

neither of which, in general, is the model internal signal $U(z)$ defined in equation 5.3, but both of whose power spectra are equal, and also equal to that of $U(z)$.

$$U_L(e^{j\theta})U_L^*(e^{j\theta}) = U_R(e^{j\theta})U_R^*(e^{j\theta}) = U(e^{j\theta})U^*(e^{j\theta}) \quad (5.29)$$

Each of the original FIR filters has been expanded into the combination of the equivalent minimum-phase filter, and an all-pass filter which provides the excess phase shift.

This expanded model is then transformed by manipulation of the blocks into the form shown in the lower half of the figure. The recursive part of the original C-ARMA can now be extracted as shown in figure 5.3. This is now seen to consist of two distinct (though non-independent) white noise sources driving two instances of the same all-pole filter to produce the signals $U_L(z)$ and $U_R(z)$. The form of this part of the C-ARMA model is identical to the Ensemble-AR system discussed in chapter 4 and the model poles may therefore be estimated by the methods discussed there.

5.3.3 Verification of Parameter Estimation Algorithm

Synthetic data was generated using a number of different C-ARMA models to assess the parameter estimation algorithm described above. In each of the figures (5.4, 5.5, 5.6) the true model and the estimated model for each of the left and right output signals is shown.

The three tests shown are of

- a system with non-minimum-phase zeros (figure 5.4),
- a system in which zeros in the FIR part cancel poles in the recursive part (figure 5.5),

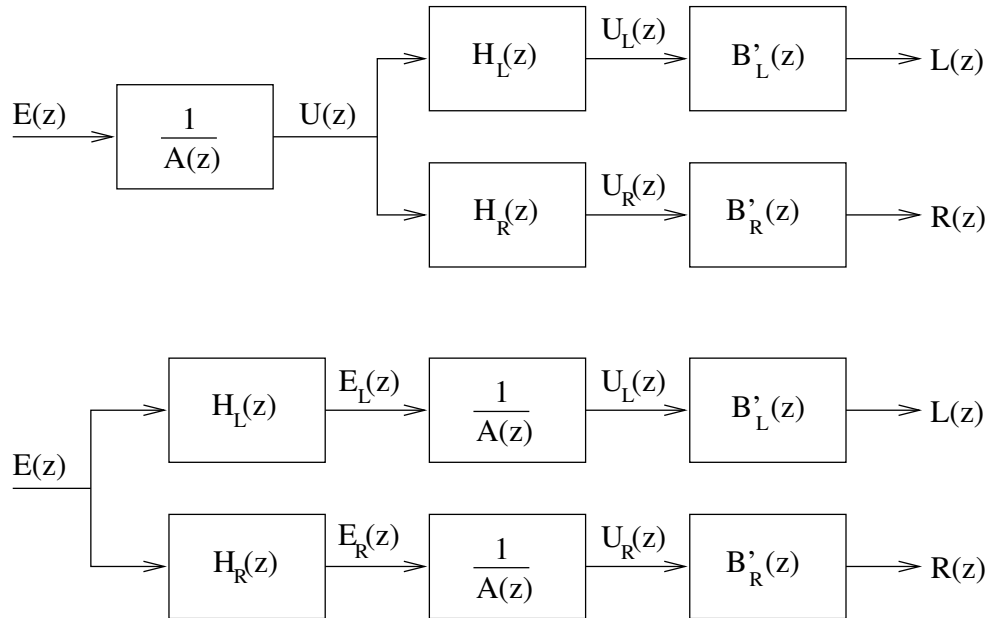


 FIGURE 5.2: *Expanded C-ARMA Model (Two Equivalent Forms)*

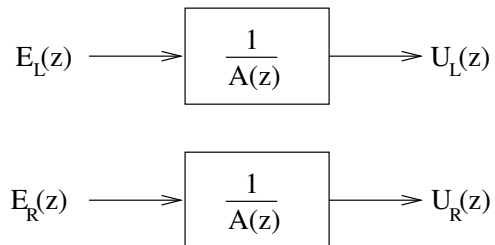


 FIGURE 5.3: *Expanded C-ARMA Model (Recursive Part)*

- a system which has zeros on the unit circle (figure 5.6).

Each of these cases has been chosen to test a particular aspect of the model estimation algorithm.

5.3.3.1 Test One — Non-Minimum-Phase Zeros

Firstly, non-minimum-phase zeros are included to test the stabilisation of the inverse system, and to verify that this does not affect determination of the system poles. Both the zeros and the poles are accurately estimated; note that in this system both channels have non-minimum-phase zeros. Also note that the poles are accurate in spite of the non-minimum-phase zeros, verifying that the transformation to a minimum-phase system does not adversely affect the estimation of the system poles.

The fact that the system has non-minimum-phase zeros in both channels is particularly interesting. This results in both $B_L(z)$ and $B_R(z)$ having roots outside the unit circle, which further implies that both inter-channel transfer functions $\frac{B_L(z)}{B_R(z)}$ and $\frac{B_R(z)}{B_L(z)}$ have poles outside the unit circle, and are therefore unstable.

5.3.3.2 Test Two — Pole-Zero Cancellation

The second case we consider is that in which poles in the source signal model are cancelled by zeros in the FIR part. All the model poles and zeros are accurately determined; note that in this case it would be impossible to estimate these cancelled poles and zeros without the presence of the second signal. Thus the multi-channel approach shows a great benefit in determining the parameters of the underlying audio signal.

5.3.3.3 Test Three — Zeros on the Unit Circle

Thirdly, zeros on the unit circle are investigated. These are a potential source of problems since the inter-channel transfer function becomes marginally stable in this case, and therefore cannot be transformed into a pair of stable inverse filters. The zeros have been estimated to be just inside the unit circle, but this has not significantly affected the accuracy with which the system poles are determined.

5.3.4 Total Least Squares Zero-Estimation

The pole-zero estimations shown in figures 5.4–5.6 are good, but the procedure described in section 5.3 has been found to be inaccurate in two particular scenar-

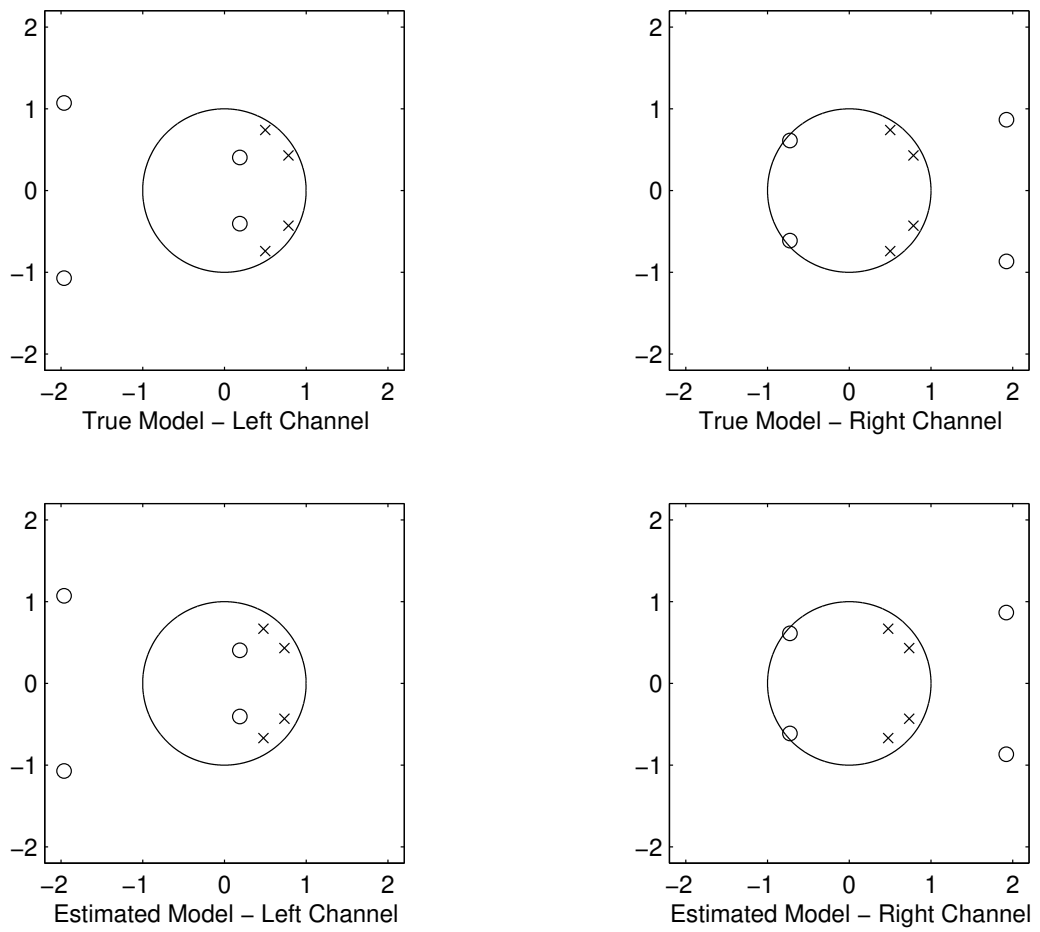


FIGURE 5.4: *Non-minimum-phase Model Estimation*

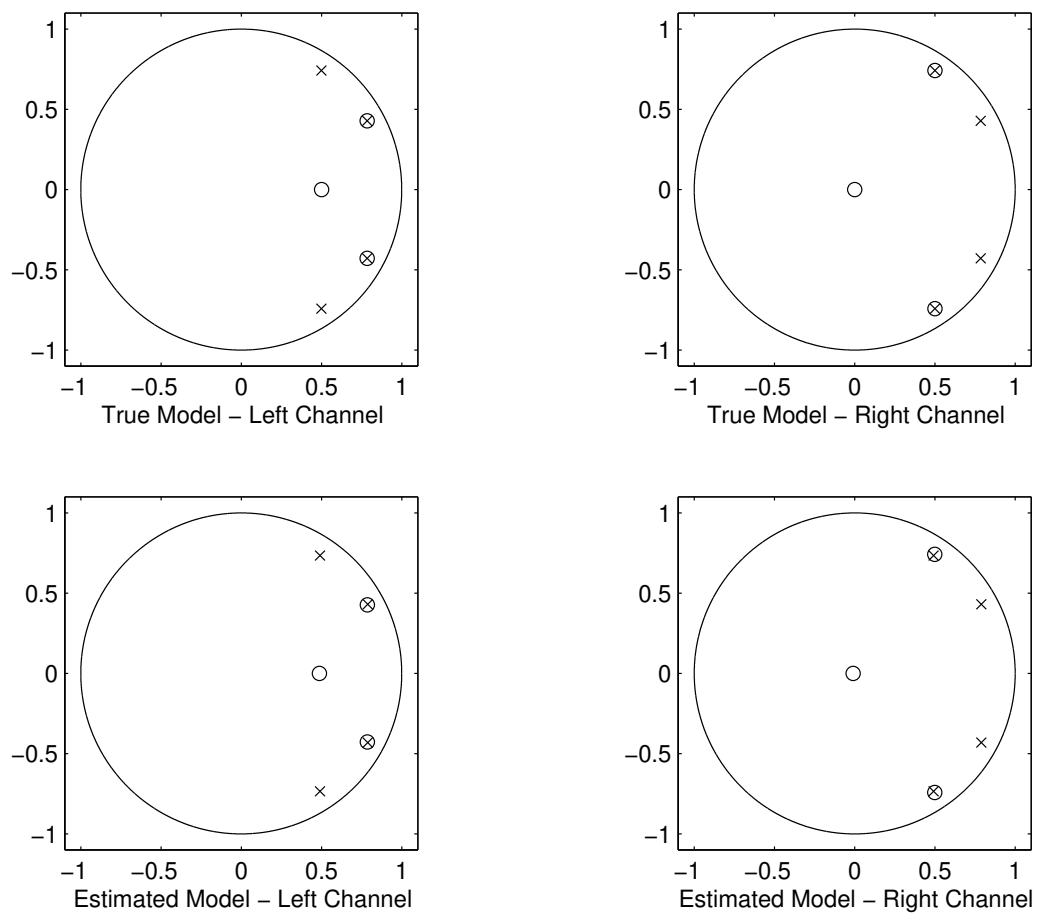


FIGURE 5.5: Pole-Zero Cancellation

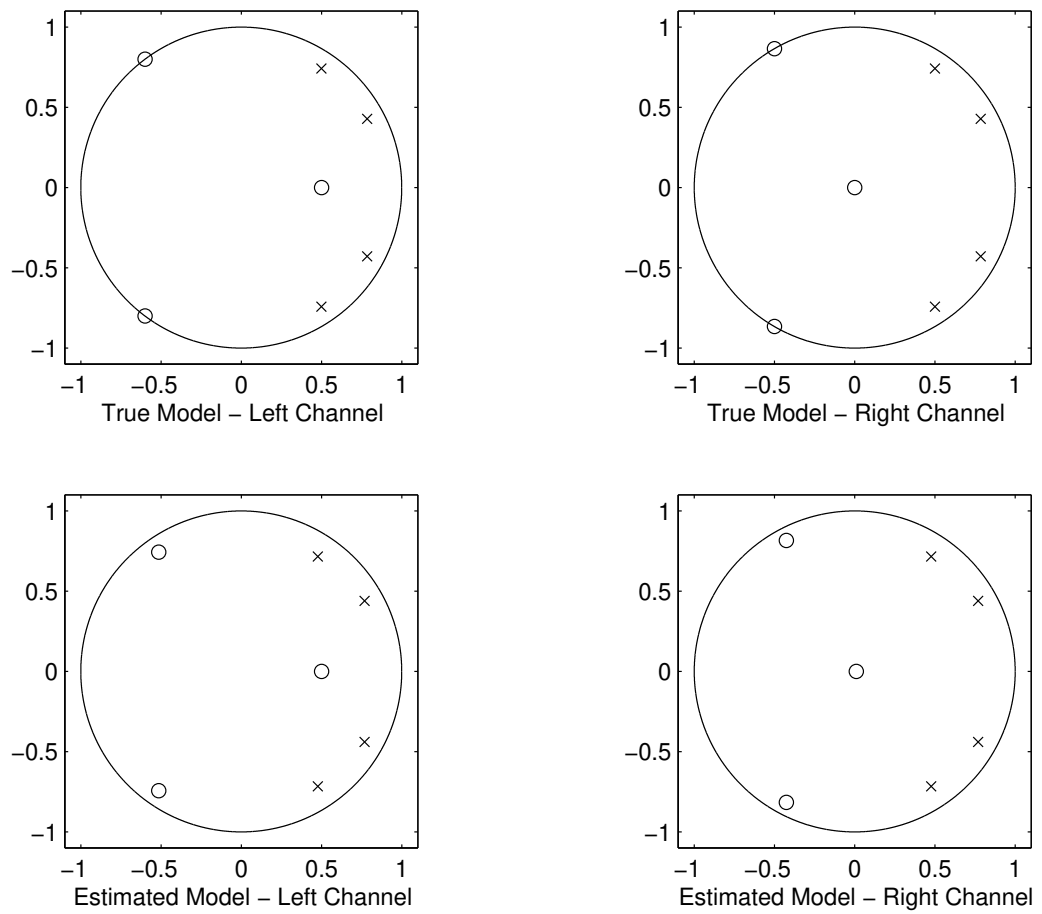


FIGURE 5.6: Zeros on Unit Circle

ios:

- when the two FIR sections have similar but non-identical zeros, and
- when the signals are contaminated by significant observation noise.

Both of these problems affect determination of the zeros more seriously than determination of the poles. It has been suggested [42] that for problems of this type the Total Least Squares (henceforth TLS) method can give superior results.

The TLS method is detailed in appendix E. It provides a solution $\underline{\mathbf{b}}_{\text{TLS}}$ to systems of the form

$$\mathbf{X}\underline{\mathbf{b}} \approx \underline{\mathbf{d}} \quad (5.30)$$

which assumes that there are errors associated with both the matrix \mathbf{X} and the vector $\underline{\mathbf{d}}$. The ordinary least squares method, by contrast, assumes that the errors are associated with $\underline{\mathbf{d}}$ alone.

We repeat equation 5.5

$$\sum_{i=0}^P \mathbf{b}_R[i] \mathbf{l}[n-i] = \sum_{i=0}^P \mathbf{b}_L[i] \mathbf{r}[n-i] \quad (5.31)$$

for clarity. Once again we constrain, with no loss of generality, $\mathbf{b}_R[0] = 1$, and because of observational noise and modelling error we replace the equality with an approximation to give

$$\mathbf{l}[n] + \sum_{i=1}^P \mathbf{b}_R[i] \mathbf{l}[n-i] \approx \sum_{i=0}^P \mathbf{b}_L[i] \mathbf{r}[n-i]. \quad (5.32)$$

Writing this as a matrix equation for a finite block of N samples we obtain

$$\underline{\mathbf{l}} + \mathbf{L}\underline{\mathbf{b}}_R \approx \mathbf{R}\underline{\mathbf{b}}_L \quad (5.33)$$

which may be rearranged thus

$$\begin{bmatrix} \mathbf{R} & -\mathbf{L} \end{bmatrix} \begin{bmatrix} \underline{\mathbf{b}}_L \\ \underline{\mathbf{b}}_R \end{bmatrix} \approx \underline{\mathbf{l}} \quad (5.34)$$

$$\mathbf{X}\underline{\mathbf{b}} \approx \underline{\mathbf{l}} \quad (5.35)$$

which is seen to be of the same form as equation 5.30. Furthermore, both the matrix \mathbf{X} and vector $\underline{\mathbf{l}}$ are subject to observation error, the primary justification

for use of the TLS algorithm. Inclusion of these errors allows restoration of the equality

$$(\mathbf{X} - \mathbf{E}) \underline{\mathbf{b}} = \underline{\mathbf{l}} - \underline{\mathbf{e}} \quad (5.36)$$

and solution by the TLS method.

The TLS solution is given *via* the singular value decomposition of the matrix

$$\mathbf{U}\Sigma\mathbf{V}^T = \left[\begin{array}{c|c} \mathbf{X} & \underline{\mathbf{l}} \end{array} \right] \quad (5.37)$$

and minimises the Frobenius norm of the error matrix

$$\mathbf{W} = \left[\begin{array}{c|c} \mathbf{E} & \underline{\mathbf{e}} \end{array} \right]. \quad (5.38)$$

The solution is given by the right-hand singular vector $\underline{\mathbf{v}}_{p+1}$ which corresponds to the smallest singular value. The final element v of $\underline{\mathbf{v}}_{p+1}$ is partitioned from the rest

$$\underline{\mathbf{v}}_{p+1} = \left[\begin{array}{c} \underline{\mathbf{v}} \\ v \end{array} \right] \quad (5.39)$$

and the solution is then given by

$$\underline{\mathbf{b}}_{\text{TLS}} = -\frac{1}{v} \underline{\mathbf{v}} \quad (5.40)$$

and is thus readily computed, though with a larger number of computations than for the ordinary least squares solution given by 5.14.

The TLS method was tested by estimating the parameters for synthetic data under a number of conditions. In each test, ten different random data sets were prepared, and the zeros estimated by both the ordinary LS and TLS methods, the results of which were then compared.

5.3.4.1 Test One — Near-Identical Zeros

Figure 5.7 shows an example where there is a pair of zeros which are similar, but not identical, in both channels. In this, and subsequent, figures the true system zeros are shown by the symbol \circ , and the estimated zeros by \times . The TLS method has been markedly more successful at correctly identifying the system zeros.

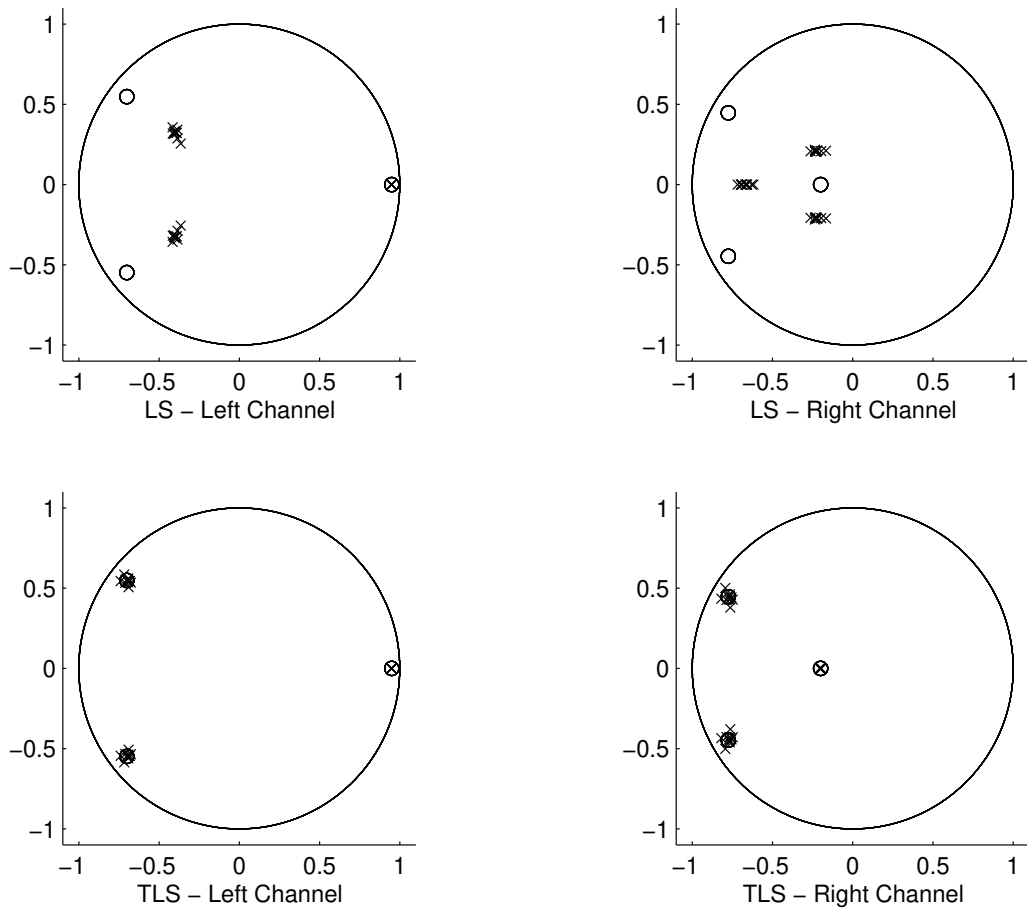


FIGURE 5.7: *Similar zeros in both channels*

5.3.4.2 Test Two — Varying SNR

Figures 5.8–5.10 show zero estimates for the same system at different signal to noise ratios of approximately 66 dB, 46 dB and 26 dB respectively.

From figure 5.8 it is seen that at 66 dB SNR both the ordinary LS and TLS methods give accurate results. As the SNR decreases the accuracy of the ordinary LS deteriorates first, as shown in figure 5.9, which is for signals of approximately 46 dB.

As the SNR decreases further, neither algorithm accurately determines the system zeros. However, whereas the LS algorithm estimates a minimum-phase system, the system estimated by the TLS method frequently includes zeros outside the unit circle. This difference is clearly seen in figure 5.10. Furthermore, the lack of clustering of the zero estimates in the TLS case implies that the model estimate is taking little regard of the signal characteristics, but rather modelling the noise component of the signal. This feature of the TLS algorithm in this application stems from the fact that the problem formulation includes many more error terms than observed sample values.

5.3.4.3 Test Three — Over-Estimated Model Order

Figure 5.11 shows the systems estimated when the model order has been over-estimated. In this case five zeros were estimated for a system with only three true zeros. The ordinary LS method puts the extra zeros close to the origin, whereas the TLS method places them close to the unit circle where they can have a large impact on the frequency response of the model. The figure shows just one representative example of the ten systems estimated for the ten original data sets.

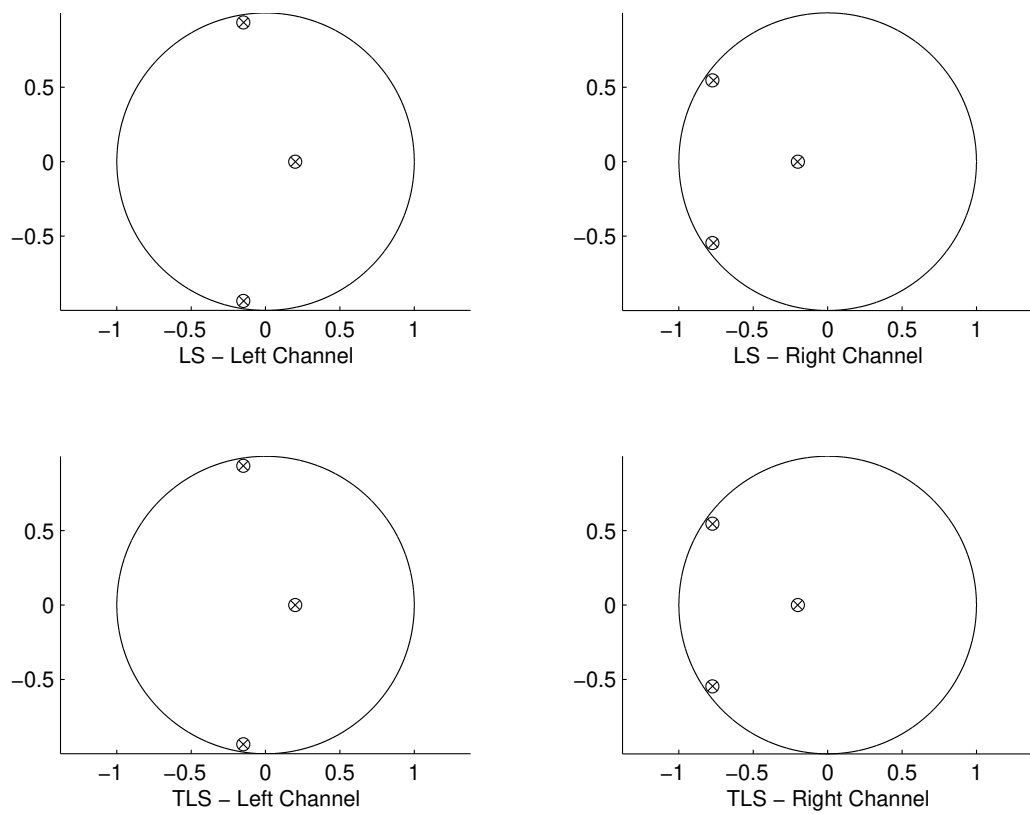


FIGURE 5.8: Zero Estimation — 66 dB SNR

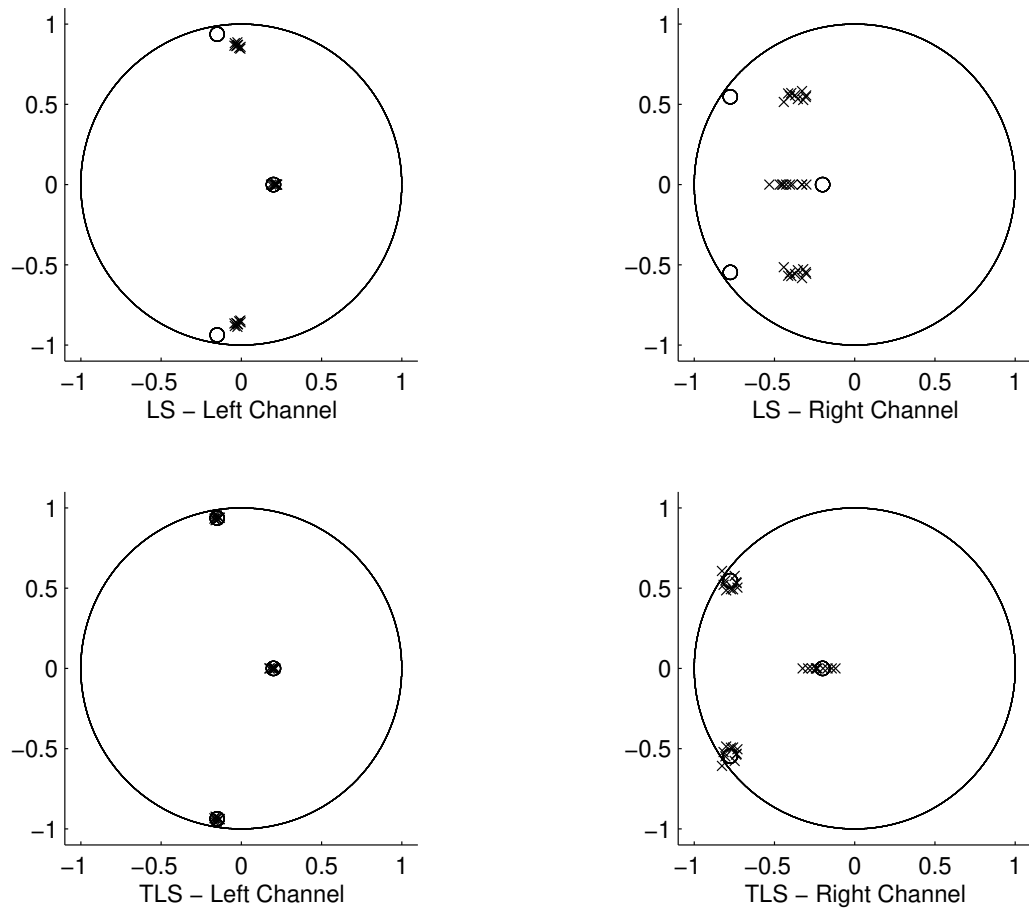


FIGURE 5.9: Zero Estimation — 46 dB SNR

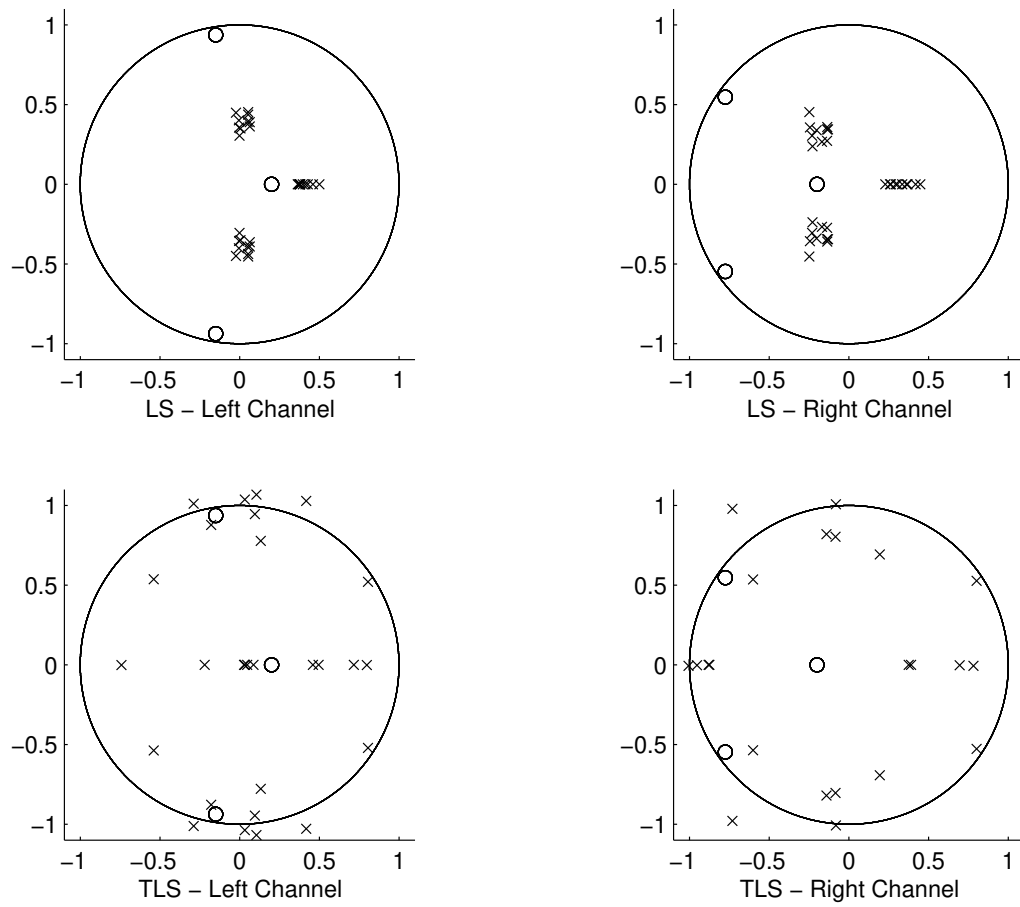


FIGURE 5.10: Zero Estimation — 26 dB SNR

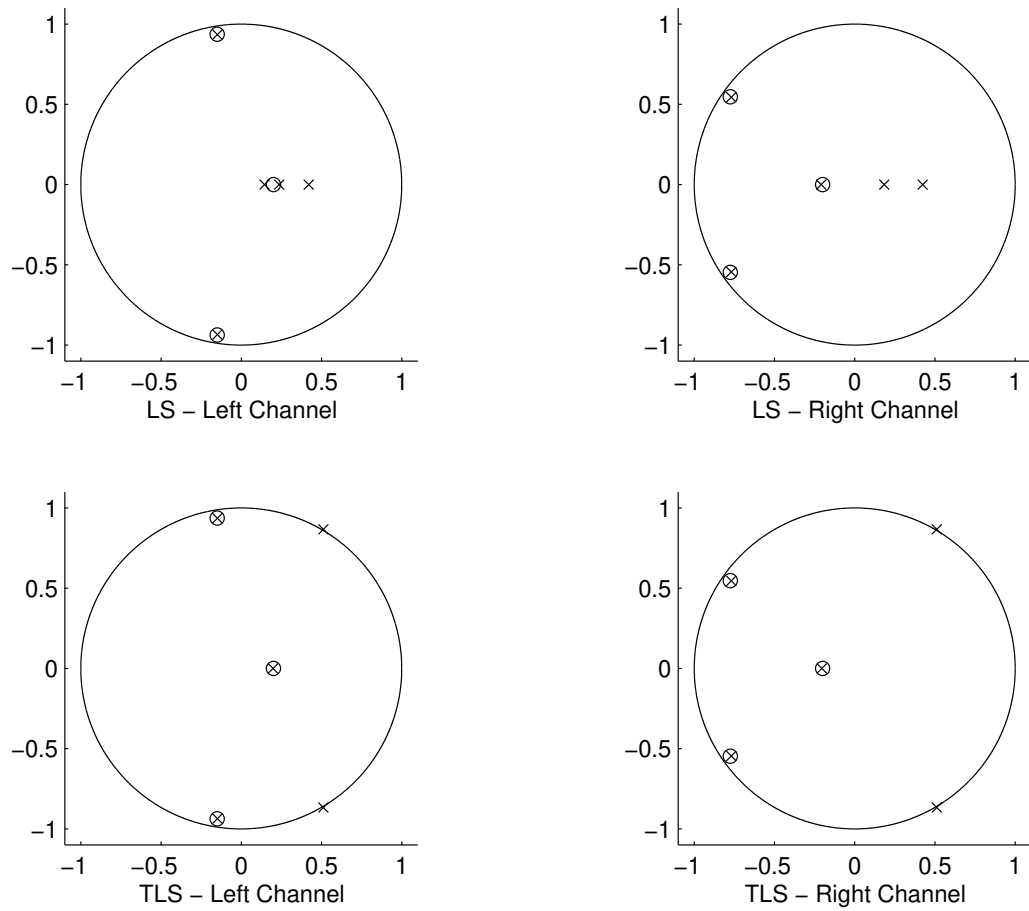


FIGURE 5.11: *Over-Estimated Model Order*

5.4 Interpolation of Missing Data

We have chosen estimation and interpolation of missing data as an illustrative application of the C-ARMA model. This problem arises, for example, in the removal of impulsive noise from audio recordings. An interpolator for signals modelled as ARMA processes has been developed by Soon Leng Ng [74], and we use this work in the form presented by Godsill and Rayner [40] as the basis for development of a C-ARMA interpolator.

We first outline the description of the method given by Godsill and Rayner for MAP interpolation of a Gaussian signal of known covariance, and the application of this to ARMA signals. We then extend the method to the joint interpolation of samples from both channels of a C-ARMA system.

5.4.1 MAP Interpolation of Gaussian Signals

Consider a frame of data \underline{x} from a Gaussian process whose covariance matrix \mathbf{R}_x is known. Some of the data samples contained within this frame are known, and some are unknown. We wish to obtain an estimate of the unknown samples.

5.4.1.1 Zero-mean Signals

We partition the data vector \underline{x} into those samples which are known (\underline{x}_k) and unknown (\underline{x}_u). We also define two matrices \mathbf{K} and \mathbf{U} which reassemble \underline{x} from the partitioned data such that

$$\underline{x} = \mathbf{K}\underline{x}_k + \mathbf{U}\underline{x}_u \quad (5.41)$$

Matrices \mathbf{U} and \mathbf{K} are complementary column-wise partitions of the identity matrix.

We may write the p.d.f. for the data vector as

$$p(\underline{x}) = p(\underline{x}_u | \underline{x}_k) p(\underline{x}_k) \quad (5.42)$$

using the probability chain rule. Rearrangement, and substitution of expression 5.41 gives

$$p(\underline{x}_u | \underline{x}_k) = \frac{p(\mathbf{K}\underline{x}_k + \mathbf{U}\underline{x}_u)}{p(\underline{x}_k)} \quad (5.43)$$

as the p.d.f. of the unknown samples, conditional on the known samples.

The general form for the zero-mean multivariate Gaussian is given in equation C.1, and is repeated here

$$p_{\mathbf{x}}(\underline{\mathbf{x}}) = \frac{1}{(2\pi)^{N/2} |\mathbf{R}_{\mathbf{x}}|^{1/2}} \exp\left(-\frac{\underline{\mathbf{x}}^T \mathbf{R}_{\mathbf{x}}^{-1} \underline{\mathbf{x}}}{2}\right) \quad (5.44)$$

for the case of zero mean, and where the vector $\underline{\mathbf{x}}$ is of length N .

Substituting 5.41 for $\underline{\mathbf{x}}$ in 5.44 gives

$$p_{\mathbf{x}}(\underline{\mathbf{x}}) = \frac{1}{(2\pi)^{N/2} |\mathbf{R}_{\mathbf{x}}|^{1/2}} \exp\left(-\frac{(\mathbf{K}_{\underline{\mathbf{x}}_k} + \mathbf{U}_{\underline{\mathbf{x}}_u})^T \mathbf{R}_{\mathbf{x}}^{-1} (\mathbf{K}_{\underline{\mathbf{x}}_k} + \mathbf{U}_{\underline{\mathbf{x}}_u})}{2}\right) \quad (5.45)$$

The MAP interpolation is given by maximisation of 5.43 with respect to the unknown samples. Since $p(\underline{\mathbf{x}}_k)$ is constant, and $\exp(\cdot)$ is a monotonically increasing function, this maximisation is directly equivalent to minimisation of the expression

$$(\mathbf{K}_{\underline{\mathbf{x}}_k} + \mathbf{U}_{\underline{\mathbf{x}}_u})^T \mathbf{R}_{\mathbf{x}}^{-1} (\mathbf{K}_{\underline{\mathbf{x}}_k} + \mathbf{U}_{\underline{\mathbf{x}}_u}) \quad (5.46)$$

again with respect to the unknown samples.

This minimisation is tractable by standard vector-matrix calculus and yields

$$\hat{\underline{\mathbf{x}}}_u = -(\mathbf{U}^T \mathbf{R}_{\mathbf{x}}^{-1} \mathbf{U})^{-1} \mathbf{U}^T \mathbf{R}_{\mathbf{x}}^{-1} \mathbf{K}_{\underline{\mathbf{x}}_k} \quad (5.47)$$

as the estimate of the unknown data samples. The interpolated data vector is reassembled by equation 5.41, substituting $\hat{\underline{\mathbf{x}}}_u$ for the unknown samples.

5.4.1.2 Non-zero Mean Signals

If the signal $\underline{\mathbf{x}}$ has non-zero mean $\underline{\mathbf{m}}_{\mathbf{x}}$ MAP interpolation is given by

$$\hat{\underline{\mathbf{x}}}_u = -(\mathbf{U}^T \mathbf{R}_{\mathbf{x}}^{-1} \mathbf{U})^{-1} \mathbf{U}^T \mathbf{R}_{\mathbf{x}}^{-1} (\mathbf{K}_{\underline{\mathbf{x}}_k} - \underline{\mathbf{m}}_{\mathbf{x}}). \quad (5.48)$$

The derivation is a simple extension to the zero-mean case.

5.4.2 Interpolation of Gaussian ARMA Signals

It was shown in section 5.3.2 that in the case of a non-minimum-phase C-ARMA system we can express the MA model sections as the combination of a minimum-phase filter and an all-pass filter. Figure 5.2 shows that the system may be re-arranged so as to separate these all-pass filters from the rest of the system. We are now able to extract two minimum-phase ARMA systems as shown in figure 5.12.

The ARMA interpolator described above is directly applicable to each of these component systems. In the following description of the algorithm we assume that the MA parameter vector has been transformed to this invertible form. We also assume that $N > 2P$ where N is the data vector length, and P is the model order.

In order to apply the MAP interpolator to an ARMA signal an expression for the covariance matrix \mathbf{R}_x is required. In section 2.5 we gave matrix expressions

$$\underline{x} = \mathbf{B}\underline{u} \quad (5.49)$$

$$\underline{e} = \mathbf{A}\underline{u} \quad (5.50)$$

for the ARMA difference equation in terms of an internal AR process $u[n]$, where matrix \mathbf{B} is defined as

$$\begin{bmatrix} \mathbf{b}_P & \mathbf{b}_{P-1} & \cdots & \mathbf{b}_0 & 0 & \cdots & \cdots & 0 \\ 0 & \mathbf{b}_P & \mathbf{b}_{P-1} & \cdots & \mathbf{b}_0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \mathbf{b}_P & \mathbf{b}_{P-1} & \cdots & \mathbf{b}_0 & 0 \\ 0 & \cdots & \cdots & 0 & \mathbf{b}_P & \mathbf{b}_{P-1} & \cdots & \mathbf{b}_0 \end{bmatrix} \quad (5.51)$$

and \mathbf{A} as

$$\begin{bmatrix} -\mathbf{a}_P & -\mathbf{a}_{P-1} & \cdots & -\mathbf{a}_1 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & -\mathbf{a}_P & -\mathbf{a}_{P-1} & \cdots & -\mathbf{a}_1 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -\mathbf{a}_P & -\mathbf{a}_{P-1} & \cdots & -\mathbf{a}_1 & 1 & 0 \\ 0 & \cdots & \cdots & 0 & -\mathbf{a}_P & -\mathbf{a}_{P-1} & \cdots & -\mathbf{a}_1 & 1 \end{bmatrix}. \quad (5.52)$$

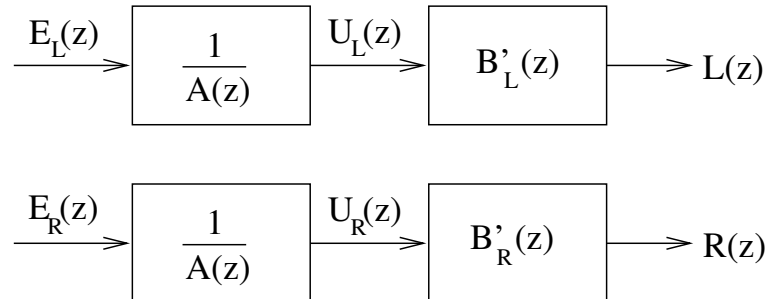


FIGURE 5.12: Two minimum-phase ARMA systems extracted from C-ARMA system

If we partition A and B column-wise to split the first P columns (subscript 0) from the rest (subscript 1), then an equivalent partitioning of \underline{u} enables us to write

$$\underline{x} = B_0 \underline{u}_0 + B_1 \underline{u}_1 \quad (5.53)$$

$$\underline{e} = A_0 \underline{u}_0 + A_1 \underline{u}_1. \quad (5.54)$$

We also define the matrix A_2 as the final P columns of A .

Since B_1 is lower triangular, has a non-zero leading diagonal (\mathbf{b}_0), and represents a minimum-phase filter, it is invertible. We may therefore re-arrange equation 5.53 as

$$\underline{u}_1 = B_1^{-1} (\underline{x} - B_0 \underline{u}_0) \quad (5.55)$$

and substitute \underline{u}_1 in equation 5.54 to give

$$\underline{e} = A_0 \underline{u}_0 + A_1 B_1^{-1} (\underline{x} - B_0 \underline{u}_0) \quad (5.56)$$

$$= A_1 B_1^{-1} \underline{x} + (A_0 - A_1 B_1^{-1} B_0) \underline{u}_0 \quad (5.57)$$

$$= F \underline{x} + G \underline{u}_0 \quad (5.58)$$

It can be shown [40] that the p.d.f. of the data is given by

$$p_x(\underline{x}) \propto \exp \left(-\frac{1}{2\sigma_e^2} (\underline{x}^T F^T (I - G(G^T G + M^{-1})^{-1} G^T) F \underline{x}) \right) \quad (5.59)$$

and, by comparison with equation 5.44, the inverse covariance matrix by

$$\sigma_e^2 R_x^{-1} = F^T \left(I - G (G^T G + M^{-1})^{-1} G^T \right) F \quad (5.60)$$

where M^{-1} is the inverse covariance matrix

$$M^{-1} = (A_2^T A_2)^R - A_0^T A_0 \quad (5.61)$$

for P samples of the AR process. The operator R reverses the rows and columns of a matrix.

The complete interpolator is formed by substitution of the inverse covariance, given by equation 5.60, into the general MAP interpolator given by equation 5.47.

5.4.3 Stereo C-ARMA Interpolator

The two-channel C-ARMA system is described by the matrix equations

$$\underline{e} = A\underline{u} \quad (5.62)$$

$$\underline{l} = B_L\underline{u} \quad (5.63)$$

$$\underline{r} = B_R\underline{u} + \underline{\epsilon} \quad (5.64)$$

where $\underline{\epsilon}$ is the modelling error described in section 5.3.

If we can find an expression for $p(\underline{l}, \underline{r})$, the joint p.d.f. of the observed data, then maximisation of this with respect to any unknown samples of both channels will give an interpolation of the stereo data, based upon this joint p.d.f.

5.4.3.1 Conditional PDF of Two-Channel Data

The first stage in calculating the joint p.d.f. is to derive an expression for the p.d.f. of the right channel data, given the left channel data, $p(\underline{r}|\underline{l})$.

By partitioning analogous to that of section 5.4.2 the latent AR process can be expressed as

$$\underline{u}_1 = B_{L1}^{-1}(\underline{l} - B_{L0}\underline{u}_0) \quad (5.65)$$

$$\underline{u}_1 = B_{R1}^{-1}(\underline{r} - \underline{\epsilon} - B_{R0}\underline{u}_0). \quad (5.66)$$

Elimination of \underline{u}_1 allows us to express $\underline{\epsilon}$ as

$$\underline{\epsilon} = \underline{r} + (B_{R1}B_{L1}^{-1}B_{L0} - B_{R0})\underline{u}_0 - B_{R1}B_{L1}^{-1}\underline{l} \quad (5.67)$$

$$= \underline{r} + G\underline{u}_0 - F\underline{l} \quad (5.68)$$

where $F = B_{R1}B_{L1}^{-1}$ and $G = (B_{R1}B_{L1}^{-1}B_{L0} - B_{R0})$.

Assuming $\underline{\epsilon}$ is distributed as a white zero-mean Gaussian, and is independent of \underline{e} we can write

$$p(\underline{r} | \underline{l}, \underline{u}_0) \propto \exp\left(-\frac{1}{2\sigma_\epsilon^2}\underline{\epsilon}^T\underline{\epsilon}\right) \quad (5.69)$$

$$p(\underline{r}, \underline{u}_0 | \underline{l}) \propto \exp\left(-\frac{1}{2\sigma_\epsilon^2}\underline{\epsilon}^T\underline{\epsilon} - \frac{1}{2\sigma_\epsilon^2}\underline{u}_0^T M^{-1}\underline{u}_0\right) \quad (5.70)$$

and substituting for $\underline{\epsilon}$ from equation 5.68

$$p(\underline{r}, \underline{u}_0 | \underline{l}) \propto \exp\left(-\frac{1}{2\sigma_\epsilon^2}(\underline{r} + G\underline{u}_0 - F\underline{l})^T(\underline{r} + G\underline{u}_0 - F\underline{l}) - \frac{1}{2\sigma_\epsilon^2}\underline{u}_0^T M^{-1}\underline{u}_0\right) \quad (5.71)$$

Marginalising \underline{u}_0 using the results from appendix C we obtain

$$p(\underline{r} | \underline{l}) \propto \exp \left(-\frac{1}{2\sigma_\epsilon^2} \left((\underline{r} - F\underline{l})^\top \left(I - G \left(G^\top G + \frac{\sigma_\epsilon^2}{\sigma_e^2} M^{-1} \right)^{-1} G^\top \right) (\underline{r} - F\underline{l}) \right) \right) \quad (5.72)$$

as the required conditional p.d.f.

5.4.3.2 PDF of Single-Channel ARMA Data

The second step is to derive the p.d.f. of the single channel \underline{l} . This is identical to the p.d.f. of the ARMA process \underline{x} (equation 5.59) and we may write immediately

$$p(\underline{l}) \propto \exp \left(-\frac{1}{2\sigma_e^2} (\underline{l}^\top D^\top (I - C(C^\top C + M^{-1})^{-1} C^\top) D \underline{l}) \right) \quad (5.73)$$

where $C = A_0 - A_1 B_{L1}^{-1} B_{L0}$ and $D = A_1 B_{L1}^{-1}$.

5.4.3.3 Stereo Data Joint PDF

We may now derive the required joint p.d.f. as

$$p(\underline{l}, \underline{r}) = p(\underline{r} | \underline{l}) p(\underline{l}) \quad (5.74)$$

by the standard rules of conditional probability.

Substituting terms into equation 5.74 from 5.72 and 5.73 gives

$$p(\underline{l}, \underline{r}) \propto \exp \left(-\frac{1}{2} \left((\underline{r} - F\underline{l})^\top R_1^{-1} (\underline{r} - F\underline{l}) + \underline{l}^\top R_2^{-1} \underline{l} \right) \right) \quad (5.75)$$

where

$$R_1^{-1} = \frac{1}{\sigma_\epsilon^2} \left(I - G \left(G^\top G + \frac{\sigma_\epsilon^2}{\sigma_e^2} M^{-1} \right)^{-1} G^\top \right) \quad (5.76)$$

$$R_2^{-1} = \frac{1}{\sigma_e^2} D^\top (I - C(C^\top C + M^{-1})^{-1} C^\top) D \quad (5.77)$$

Some algebraic manipulation allows us to write the joint p.d.f. in the standard form

$$p(\underline{l}, \underline{r}) \propto \exp \left(-\frac{1}{2} \begin{bmatrix} \underline{l}^\top & \underline{r}^\top \end{bmatrix} \begin{bmatrix} F^\top R_1^{-1} F + R_2^{-1} & -F^\top R_1^{-1} \\ -R_1^{-1} F & R_1^{-1} \end{bmatrix} \begin{bmatrix} \underline{l} \\ \underline{r} \end{bmatrix} \right) \quad (5.78)$$

and defining the combined data vector \underline{d} gives

$$p(\underline{l}, \underline{r}) \propto \exp \left(-\frac{1}{2} \underline{d}^\top R_d^{-1} \underline{d} \right) \quad (5.79)$$

where

$$\underline{d} = \begin{bmatrix} \underline{l} \\ \underline{r} \end{bmatrix} \quad (5.80)$$

$$\mathbf{R}_d^{-1} = \begin{bmatrix} \mathbf{F}^T \mathbf{R}_1^{-1} \mathbf{F} \underline{l} + \mathbf{R}_2^{-1} & -\mathbf{F}^T \mathbf{R}_1^{-1} \\ -\mathbf{R}_1^{-1} \mathbf{F} & \mathbf{R}_1^{-1} \end{bmatrix} \quad (5.81)$$

This form may be used directly in the MAP interpolator (equation 5.47) to calculate joint estimates of both channels of data.

5.5 Interpolation Tests

5.5.1 Synthetic Data

Figure G.9 shows independent ARMA interpolations of a two-channel C-ARMA signal, using the method described in section 5.4.2. The interpolated signal is smooth and continuous with the original, but the interpolated section deviates significantly from the original.

Figure G.10 shows a two-channel C-ARMA interpolation made by maximisation of the joint p.d.f., equation 5.79. In this case the interpolations are seen to be a close match to the original data. In the case where there is no overlap of the interpolated regions the interpolated signal is indistinguishable (by eye, from such a figure) from the original.

5.5.2 Audio Data

The performance of the interpolation algorithm on real stereo audio data is disappointing, and it is unclear why this is so. Excerpts from modern CD recordings were used as test material, and attempts were made to estimate model parameters, and perform interpolations of the data using the various techniques developed in previous sections.

The parameter estimation procedure seems to work adequately; a pole-zero plot for an order 10 model generated in this way is shown in figure G.11. As expected there are non-minimum-phase zeros in the estimated model; note that, in this case, they occur only in the right channel. The estimated poles are all inside the unit circle.

The upper part of figure G.12 shows the stereo signal in question, and the mod-

elling error, ϵ , associated with this signal and the estimated parameters. The modelling errors are seen to be small, and analysis has shown that they are also substantially white. In addition, the model excitation signal \underline{e} was found to be many orders of magnitude smaller than \underline{u} , the internal signal, and again substantially white. The fact that the excitation and modelling error are both small and white implies that significant information regarding the structure of the signal is being carried by the model parameters.

Using the estimated model parameters to perform a single-channel ARMA interpolation yields the result shown in the lower half of figure G.12. The result is in qualitative agreement with the corresponding result for synthetic data, shown in figure G.9.

The interpolation deviates in detail from the original signal, but is smooth and continuous with it, and would undoubtedly be useful in many contexts as it retains much of the character of the original. This result further confirms that the model parameters are an accurate representation of the signal structure. Furthermore, this interpolation of 40 samples has been achieved with a model of order 10; this is a significantly lower order than has been found necessary for similar interpolations, assuming an AR model [101].

Figure G.13 shows a joint interpolation of both channels. In this case the interpolants do not follow the original signal in character. This is surprising given the previous results, and given that a similar interpolation of synthetic data (figure G.10) produced excellent results.

The reason for this discrepancy is not clear, though it is possible that the assumptions regarding whiteness and independence of \underline{e} and $\underline{\epsilon}$ are not valid in practice. Given that the result with synthetic data is good, and that there is little, qualitatively, to distinguish the synthetic and real data sets, this result is surprising and worthy of further investigation.

5.6 Conclusions

We have devised a new Coupled-ARMA model for two-channel signals which comprises a recursive part, common to both channels, and two moving-average sections, a separate one for each channel.

We have investigated methods for efficient determination of the model parameters,

based on the methods of Least Squares, and have also investigated the Total Least Squares method for this application. The TLS method was found to be more accurate under some circumstances.

Methods for calculating MAP interpolations of the C-ARMA data were investigated. An existing interpolator for ARMA signals was found to be directly applicable, and to perform satisfactorily in many circumstances.

An algorithm for making joint interpolations of both channels was derived. This was found to work excellently on synthetic data sets, but its performance on stereo audio data was lack-lustre.

Channel Synchronisation

6.1	Introduction	117
6.1.1	Problem Statement	119
6.2	Adaptive Filtering Method	121
6.2.1	Test on Real Data	122
6.3	Correlation Method	123
6.3.1	Resampling of a Windowed Signal	123
6.3.2	Cross-Correlation of Oversampled Signals	126
6.4	Model-Enhanced Correlation Method	127
6.5	Statistical Method	129
6.5.1	Offset Likelihood Function	129
6.6	Bayesian Formulation	131
6.7	Models and Priors for the Offset	131
6.7.1	AR plus Sinusoids	132
6.7.2	Parameter Estimation	134
6.7.3	Test Results	134
6.7.4	Alternative Offset Models and Priors	135
6.8	Offset Posterior PDF	137
6.8.1	MAP Offset Estimate	137
6.8.2	Joint estimate of Model Parameters and Offset	138

6.9	Tests of Model-Based Bayesian Estimator	139
6.9.1	AR+Sinusoidal Prior	139
6.9.2	Differential Smoothness Prior	139
6.10	Audio Demonstration	140
6.11	Conclusions	140

Channel Synchronisation

MULTI-CHANNEL audio work usually requires that the signals under examination are accurately aligned with other, sharing a common time origin. This is, however, difficult to achieve in practice; multiple copies of an audio tape or gramophone record will never play in exact synchronisation, and similarly it cannot be generally assumed that the recording device ran at a constant speed.

In this chapter we examine the problem of realigning audio signals under the condition of a varying time offset, and in the case where different media have imparted different frequency response anomalies on them.

6.1 Introduction

Consider a continuous-time band-limited signal $u(t)$. By processes of recording and transcription this signal is transformed into two observed signals

$$x_1(t) = h_1 \star u(f_1(t)) \tag{6.1}$$

$$x_2(t) = h_2 \star u(f_2(t)) \tag{6.2}$$

where h_1 and h_2 represent a pair of linear filters, and \star represents the convolution operator. The time axis functions f_1 and f_2 represent two “warpings” of the true time axis t .

The problem is to determine the mapping between $f_1(t)$ and $f_2(t)$ given the

pair of observed signals, x_1 and x_2 . We are interested in restoring the relative synchronisation of the signals, rather than estimating the original time axis t . The estimation of the original time axis t from a single observed signal x_1 has been explored by Godsill [35, 36] in the correction of pitch variation defects.

If we define

$$f_1(t) = t + \tau_1(t) \quad (6.3)$$

$$f_2(t) = t + \tau_2(t) \quad (6.4)$$

then from equation 6.1 we obtain

$$x_1(t - \tau_1(t) + \tau_2(t)) = h_1 \star u(t + \tau_2(t)) \quad (6.5)$$

$$= h_1 \star h_2^{-1} \star x_2(t). \quad (6.6)$$

If we now define the time *offset* $s(t) = \tau_1(t) - \tau_2(t)$ then we obtain

$$x_1(t - s) = h \star x_2(t). \quad (6.7)$$

where $h = h_1 \star h_2^{-1}$.

This analysis does not take account of the effects of the time axis warping on the filters h_1 and h_2 . If, however, the warping is slight, and the frequency responses not too drastic then this becomes an insignificant problem.

In practice the predominant feature of the filters is likely to be low-pass filtering, caused by ageing of tape, mechanical wear of gramophone records, RC filtering of analogue signals by cable and other equipment, and other similar mechanisms. The vast majority of such systems are expected to be zero-phase (in the case of mechanical processes) or minimum-phase (for the electrical mechanisms). The low-frequency phase response will, in this case, be constant and close to zero.

The effects of slight time warping on such filters will be a small modulation of the cut-off frequency, and possibly the introduction of low-level non-linear distortion artifacts. The exact effects are difficult to predict as the filtering may result from processes or transfers that occurred before, after, or even simultaneously with the processes that produce the time offset.

The non-uniform time axis will affect the audio signal (as distinct from the filter responses) in a similar way to phase modulation of the original signal. Phase modulation produces sidebands on the tonal components of the signal, whose width depends on the frequency and depth of the modulation.

6.1.1 Problem Statement

The problem we wish to solve, therefore, is to determine the instantaneous time offset between a pair of signals which represent the same musical information; subsequently we wish to resample the signals such that they are accurately synchronised. In addition to being offset in time, the signals may have undergone different linear or non-linear operations, but these and the time-axis warpings are assumed sufficiently small that the musical information is substantially intact.

6.1.1.1 Example: Shift between two copies of the same recording

If multiple copies of a single recording are available (there are many examples on the demonstration CD) then we would usually wish to synchronise them for purposes of multi-channel signal processing. Mechanical means ensure that the playback speed remains approximately constant, and this may be enhanced by electronic feedback control of the capstan or turntable motor to achieve a long-term speed stability of typically 100 p.p.m.¹

To ensure that a substantial section of music remains in synchronisation to within one sample period clearly requires a much higher degree of speed stability than is available by these mechanical and electronic means. For example, a side of an LP plays for approximately 30 minutes; one sample period over this time span represents a speed stability of 0.001 p.p.m. at the standard sample rate of 44.1 kHz.

Furthermore, playback speed fluctuations (*wow*, [8]), disc eccentricity, uneven tape tension and stretch, and so on result in a time offset between the channels which varies in an unpredictable manner.

Finally, there are great problems to be expected if we attempt to start a pair of tape players or turntables in exact synchronisation. This may be circumvented by loading the signals into a computer-based audio editor, whereupon a single, well-defined event may be aligned with good accuracy by visual inspection of the waveforms. The signals may then be replayed independently or together with complete confidence that the data will be replayed identically every time.

This procedure does not, however, solve the time-varying offset problem between multiple sources, and it is to this that we presently turn our attention.

¹Parts per million.

6.1.1.2 Example: Inter-channel shift on a single disc

It was shown in chapter 4 that it is possible to improve the restoration of a monophonic gramophone disc by extracting two signals from it, one from each groove wall. Similarly it is possible to perform multi-channel processes on single-channel tape recordings using a multi-track head, whose total span is the same as the span of the original single-channel head.

Time-shifts can exist between multiple signals extracted thus from a single carrier. The following are three commonly encountered mechanisms by which this can occur.

- The cutting stylus used to make a master for mass duplication of gramophone discs usually moves radially across the disc. By contrast, the replay arm of most high-quality transcription turntables moves in a circular arc about its pivot. Thus the sides of the replay stylus make contact with the disc at a different pair of points from the sides of the cutting stylus. This results in an offset between the channels on playback that varies across the disc. A thorough analysis of the geometry of this problem is given in [107].
- Many discs made up to about 1940 were cut on an EMI system in which the cutting stylus was mounted on a round shank. Careless alignment, and rotation of this shank in its mounting while cutting is in progress both cause groove wall offset on such discs [55].
- Inter-channel offset occurs on magnetic tape recordings as a result of inaccurate head alignment. Ideally the head gap is perpendicular to the direction of travel of the tape, and identical for both the record and replay heads. Poor alignment or inadequate mechanical integrity of either the recording or replay tape machine results in a time shift between the channels, which again can vary with time. This mechanism also results in significant frequency-response anomalies due to the effect of the finite head width.

A commercial device [18] exists which attempts to re-align two-channel signals under these conditions of a slight, slowly-varying time-offset, and which also allows realignment of stereo signals for restoration of the stereophonic illusion and improvement of mono-compatibility.

6.2 Adaptive Filtering Method

The form of this problem suggests that standard system identification procedures may be appropriate. Essentially we have a pair of signals which are related by an unknown, slowly-varying transfer function. Knowledge of this transfer function would allow us to compensate for the time offset between the channels as we require.

Let us apply an FIR filter with impulse response $\mathbf{b}[m]$ to $x_2[n]$, such that $\mathbf{b}[m] \star x_2[n] \approx x_1[n]$ where \star represents the discrete convolution operator. If the approximation is close then the amplitude response of \mathbf{b} will approach the amplitude response of \mathbf{h} , and the phase response of \mathbf{b} will be composed of the phase shift associated with the time-shift s , plus the phase response of filter \mathbf{h} . This system is shown diagrammatically in figure 6.1.

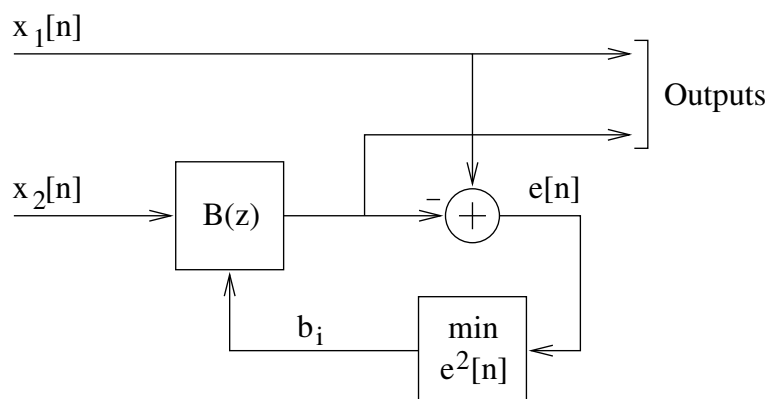


FIGURE 6.1: *Channel Synchroniser based on Adaptive Filtering*

The standard block-LMS algorithm [106] allows efficient calculation of this filter and allows tracking of a slowly-varying time offset. Furthermore, since the FIR filter may incorporate an arbitrary gain the method is robust to level differences between the two channels. Haykin [44] gives a full analysis of the convergence and tracking properties of this algorithm.

If we require simply to shift the signals into alignment, without equalising frequency response anomalies, then we may determine the low frequency group delay of \mathbf{b} from the filter phase response and shift signal x_2 by this amount. Note that the group delay may be a non-integer number of samples; the mechanism for shifting a sampled signal by fractional parts of the sample period is closely related to

sample rate conversion. Appendix F outlines the method.

6.2.1 Test on Real Data

This adaptive filtering algorithm was used to estimate the variation of time offset between [9] and [10] throughout the extracts.

The starts of the two recordings were aligned as accurately as possible using the SADiE digital audio workstation [94]. This was accomplished by examining the waveform and aligning by eye one musical event that was clearly visible in the waveform near the start of the extract. The remaining offset was then estimated independently for the two groove walls, and the results are plotted in figure 6.2.

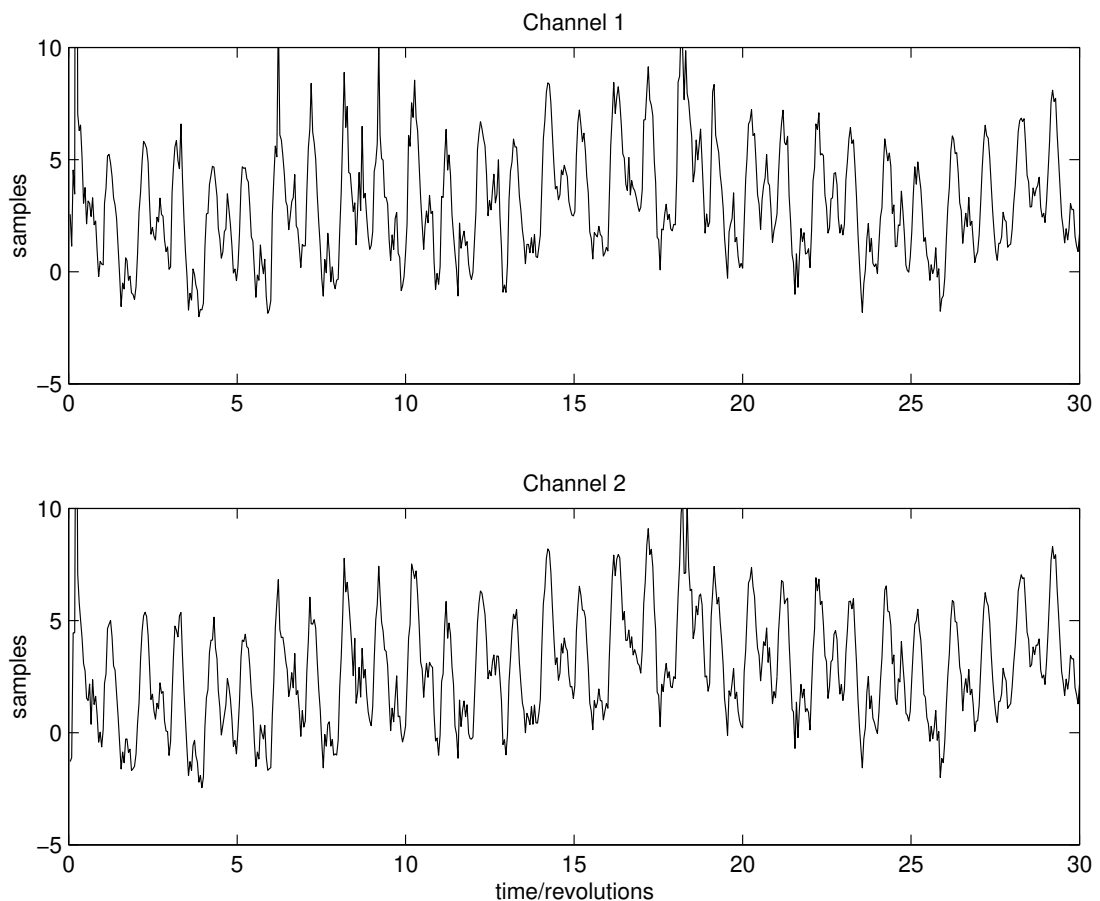


FIGURE 6.2: *Inter-channel offset estimated by the adaptive filtering method*

The estimated time offset shows a long-term drift, plus an oscillatory component at the revolution rate of the disc. The varying speed discrepancy is well below

audibility when the discs are played separately, but is obvious (as time-varying comb filtering) when they are summed, as demonstrated by track [\[11\]](#) on the demonstration CD.

The estimated offset shows a number of outliers, as well as its low-frequency variations. These outliers are due to inadequacies in the procedure used to estimate the low-frequency group delay of the filter.

6.3 Correlation Method

If we know the cross-correlation function $\mathcal{R}_{x_1x_2}(s)$ of $x_1(t)$ and $x_2(t)$, then the position of the maximum of the cross-correlation as a function of the lag s can be taken as an estimate of the time-shift between the pair of signals. The cross-correlation may be readily estimated from the data, and so this forms the basis of a detector for the inter-channel shift.

In this most basic detector the estimated time offset will always be an integer number of samples. However, bearing in mind that the origin of the time shift may operate in the continuous-time domain, we would expect to observe, at any given sample point, a non-integer shift between the channels.

It is not clear that an interpolation of the estimated cross-correlation gives a valid offset estimate of sub-sample accuracy, since the sampled signal simply does not exist between the sample points. Furthermore, the cross-correlation of random, band-limited signals $x(t)$, $y(t)$, given by

$$\mathcal{R}_{xy}(s) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} x(t) y(t-s) dt \quad (6.8)$$

is non-trivially related to the discrete cross-correlation function

$$\mathcal{R}_{xy}[k] = \lim_{N \rightarrow \infty} \frac{1}{2N} \sum_{n=-N}^{N-1} x[n] y[n-k]. \quad (6.9)$$

In particular, the latter is not a straightforward sampled form of the former.

6.3.1 Resampling of a Windowed Signal

Let us treat a given, finite block of data from a single channel as a known, isolated, set of constant data points, as opposed to treating them as a finite-length observation of a random process which is infinite in duration. We may

then justify interpolation of the cross-correlation function to obtain sub-sample accuracy by the following arguments.

Consider the set of real samples $x[nT]$, where T is the sample period. If we assume the signal is zero outside the window $0 \leq n < N$, then we may construct a real continuous-time interpolation $x(t)$ from these samples. Furthermore, if we restrict our consideration to signals $x(t)$ of bandwidth $B < \frac{1}{2T}$ then the sampling theorem states that the continuous-time signal $x(t)$ is uniquely defined by the samples $x[nT]$. For simplicity we limit the discussion to baseband signals, although this restriction is not necessary.

Now that we have reconstructed a continuous-time signal we may resample it at whatever rate we choose. In doing so we retain all of its information provided that the sampling theorem is not contravened. In particular, we are guaranteed to satisfy this criterion if we raise (not lower) the sampling rate, compared with the original rate $1/T$. Note, however, that the oversampled signal will not, in general, be windowed in the same way as the original finite data block.

Suppose that we choose to sample at a rate $\frac{P}{T}$ which is some integer factor P greater than the original rate. That is, we sample $x(t)$ and obtain the set of oversampled data points $x^P[nT^P]$, where the notation T^P is used to denote $\frac{T}{P}$, the new sampling period. This set represents the signal in a discrete-time form suitable for numeric processing, but measured on a finer time-scale than before.

It can be shown that the samples $x^P[nT^P]$ may be calculated from the original samples by the application of the linear filtering operation

$$x^P[n] = \sum_{i=-\infty}^{+\infty} h_i x^{P_0}[(n-i)T^P] \quad (6.10)$$

where $x^{P_0}[nT^P]$ has $P-1$ zero samples interposed between each of the original samples. We may thus calculate the oversampled signal directly from the original samples without reconstructing the continuous-time signal $x(t)$.

Figure 6.3 shows diagrammatically the relationships between the continuous-time signal $x(t)$, the sampled signal $x[nT]$, the oversampled signal $x^P[nT^P]$ and the zero-interleaved oversampled signal $x^{P_0}[nT^P]$.

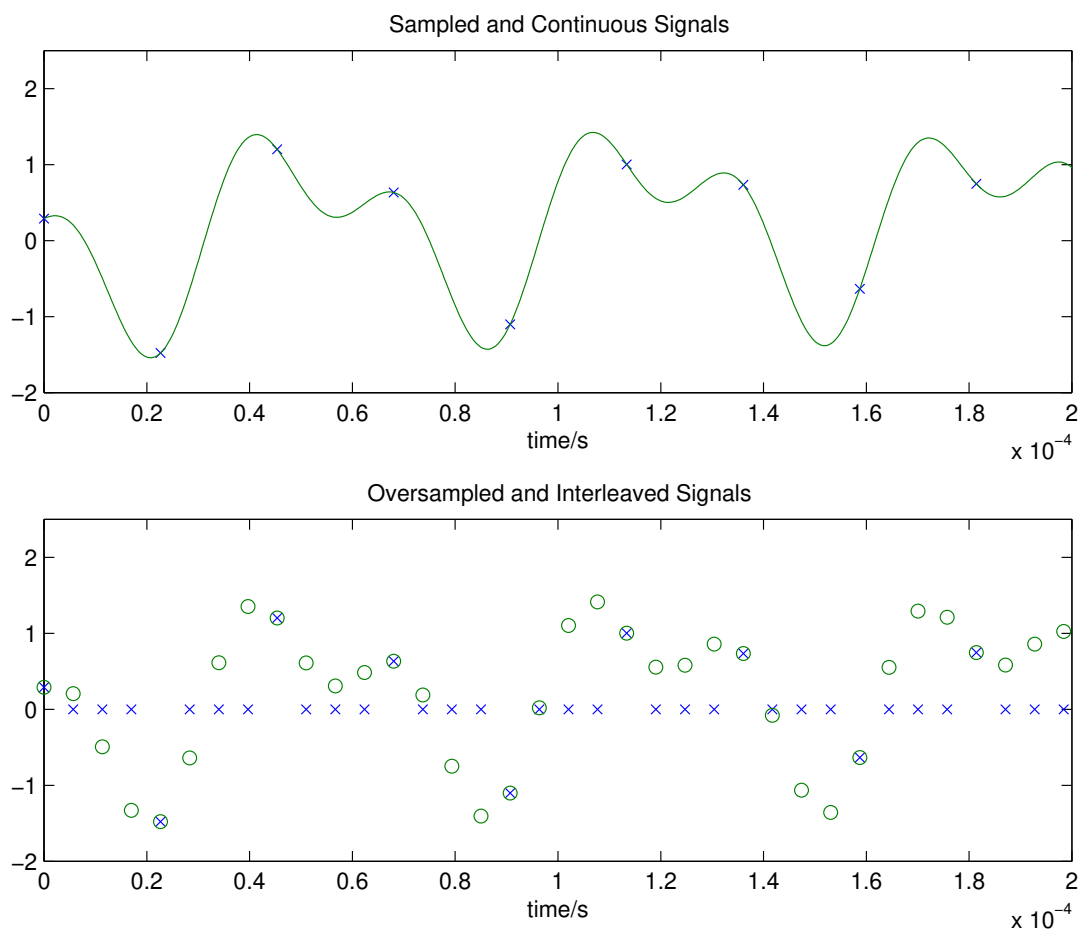


FIGURE 6.3: *Relationship of Sampled and Oversampled Signals*

6.3.2 Cross-Correlation of Oversampled Signals

We now turn our attention to the estimation of the cross-correlation of a pair of signals, $x_1^p[n]$ and $x_2^p[n]$, which have been oversampled in this manner. The cross-correlation function is defined as

$$\mathcal{R}_{x_1 x_2}[k] = \lim_{N \rightarrow \infty} \frac{1}{2N} \sum_{n=-N}^{N-1} x_1[nT] x_2[(n-k)T]. \quad (6.11)$$

If we substitute the oversampled signals we obtain

$$\mathcal{R}_{x_1 x_2}^p[k] = \lim_{N \rightarrow \infty} \frac{1}{2N} \sum_{n=-N}^{N-1} x_1^p[nT^p] x_2^p[(n-k)T^p] \quad (6.12)$$

and expressing each as a filter applied to the original samples gives

$$\begin{aligned} \mathcal{R}_{x_1 x_2}^p[k] = \\ \lim_{N \rightarrow \infty} \frac{1}{2N} \sum_{n=-N}^{N-1} \left(\sum_{i=-\infty}^{+\infty} h_i x_1^{p_0}[(n-i)T^p] \sum_{j=-\infty}^{+\infty} h_j x_2^{p_0}[(n-k-j)T^p] \right). \end{aligned} \quad (6.13)$$

We may rearrange the order of the summations, to express $\mathcal{R}_{x_1 x_2}^p[k]$ as a filter applied to the zero-interleaved signals thus

$$\begin{aligned} \mathcal{R}_{x_1 x_2}^p[k] = \\ \lim_{N \rightarrow \infty} \sum_{i=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} \left(h_i h_j \frac{1}{2N} \sum_{n=-N}^{N-1} x_1^{p_0}[(n-i)T^p] x_2^{p_0}[(n-k-j)T^p] \right) \end{aligned} \quad (6.14)$$

and hence as a filter applied to the cross-correlation of those signals

$$\mathcal{R}_{x_1 x_2}^p[k] = \sum_{i=-\infty}^{+\infty} \left(h_i \sum_{j=-\infty}^{+\infty} (h_j \mathcal{R}_{x_1 x_2}^{p_0}[k+j-i]) \right) \quad (6.15)$$

where

$$\mathcal{R}_{x_1 x_2}^{p_0}[k] = \begin{cases} \mathcal{R}_{x_1 x_2}[k/P], & \text{if } k \pmod{P} = 0 \\ 0, & \text{otherwise.} \end{cases} \quad (6.16)$$

It is clear that the non-zero values of the function $\mathcal{R}_{x_1 x_2}^{p_0}[k]$ are identical to the cross-correlation estimates of the original sampled signals $\mathcal{R}_{x_1 x_2}[k/P]$. We have now, therefore, expressed the cross-correlation estimate of the oversampled signals in terms of the cross-correlation estimate of the original signal samples.

It can be seen from equation 6.15 that the cross-correlation of the oversampled signal may be obtained from the cross-correlation of the original signal by filtering it *twice*, once with the oversampling filter h_i , and once with the same filter reversed h_{-i} .

The above analysis shows that bandlimited interpolation of the cross-correlation function is justifiable from signal processing theory. In order to perform this calculation in a practical system it will be necessary to further window the infinite summations, and the degree to which this is done will affect the accuracy of the interpolation, and hence its ability to accurately measure the inter-channel time offset.

In practice it has been found that even relatively short (for example, 7–11 taps) interpolating filters are useful in estimating the shift to sub-sample resolution. If filters of higher order are chosen then consideration should be given to computationally efficient FFT-based methods for performing the convolutions.

6.4 Model-Enhanced Correlation Method

Recall the Ensemble-AR model of chapter 4 where we had an ensemble of white excitation sequences driving an all-pole filter to give multiple observed signals. If we treat x_1 and x_2 as the outputs of such a system then we can determine the time-offset in the excitation domain, instead of in the signal domain.

This has the advantage that the estimated excitation is approximately white, due to the whitening effect of the inverse model filter. The cross-correlation of the excitation will more closely approximate a $\delta(\cdot)$ function than will the cross-correlation of the original signals, and this enables a more accurate estimation of the inter-channel shift.

The comparison between the signal cross-correlation and excitation cross-correlation is shown in figure 6.4. A block of 1024 signal samples was taken from each of [9] and [10]. The cross-correlation between these signals is plotted in the upper part of the figure.

An E-AR model of order 10 was estimated from this data and the excitation sequences corresponding to each of the original sources were calculated. The cross-correlation of these excitation sequences is plotted in the lower half of the figure. The position of the peak in the lower figure gives a more precise indication

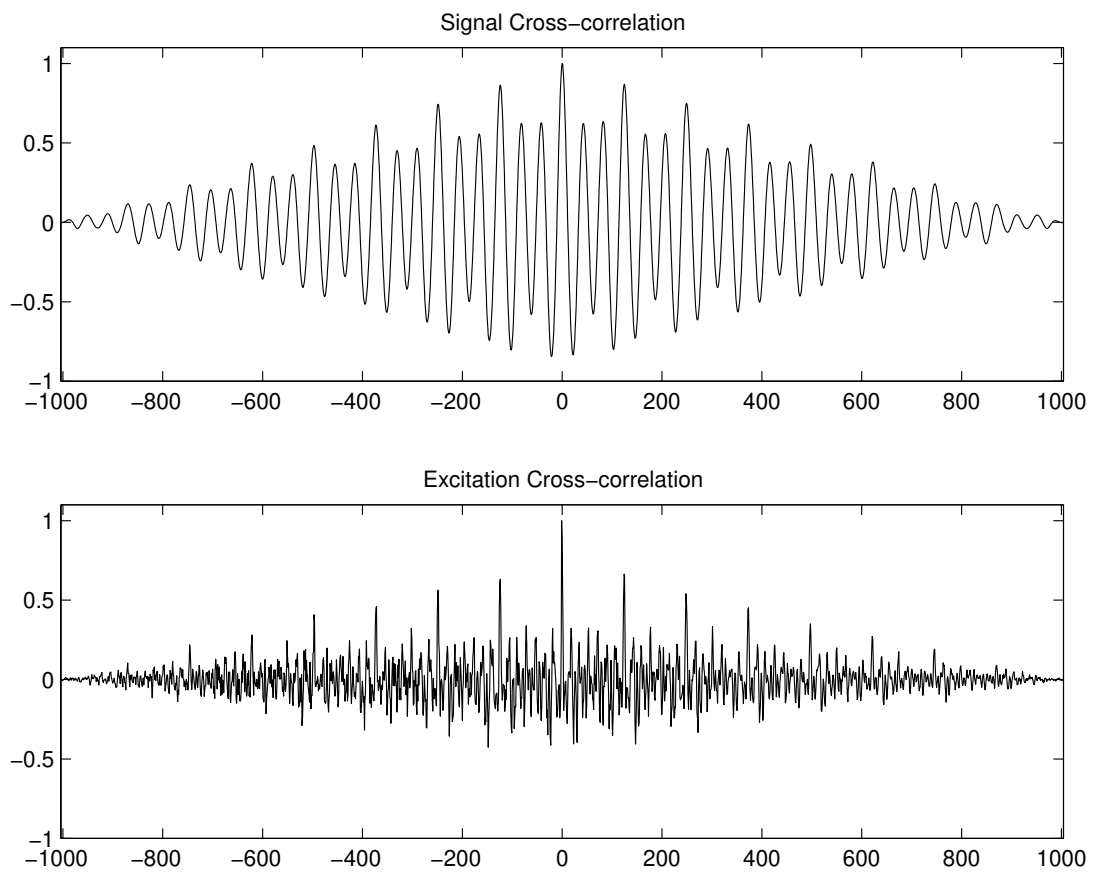


FIGURE 6.4: *Excitation and signal cross-correlations*

of the inter-channel offset than more rounded peak in the upper half of the figure.

6.5 Statistical Method

The offset fluctuation with time shows great structure, and this provides the motivation for investigating a model-based statistical method for estimating the offset. Given two signals \underline{x}_1 and \underline{x}_2 which are similar, but offset in time by s , we can derive an expression for the likelihood $p(\underline{x}_2|\underline{x}_1, s)$ and hence derive the estimate s_{ML} for the time offset.

Furthermore we can put a prior on s , based on some model of the offset variation, and thence calculate estimates for the time offset within a fully Bayesian framework. The intention is that the model-based prior will improve the resilience of the algorithm to noise in the original signals by the rigorous incorporation of qualitative prior information about the expected offset variations.

These methods have strong parallels with work by Godsill [35, 39] in the estimation of pitch fluctuations in a single recording. In the present study, however, we are concerned not with the speed fluctuations of a single recording, but rather the speed differences between a pair of transcriptions of otherwise identical nature.

6.5.1 Offset Likelihood Function

Of the two observed signals, let us arbitrarily treat $x_1(t)$ as the reference signal, and $x_2(t)$ as the second signal whose time offset $s(t)$ from x_1 we wish to determine.

We may model the second signal $x_2(t)$ as

$$x_2(t) = \alpha x_1(t - s(t)) + e(t) \quad (6.17)$$

where α compensates for any amplitude mismatch between the channels.

Let us assume, as before, that the offset varies sufficiently slowly and smoothly that we may treat frames of N consecutive signal samples as each having a constant offset². By doing this, we may replace the function $s(t)$ with the constant s in our consideration of a single data block. We may now write a discrete-time equivalent of equation 6.17 as

$$x_2[n] = \alpha x_1(nT - s) + e[n] \quad (6.18)$$

²This assumption seems justifiable given the mechanical mechanisms by which the offset originates.

where $\frac{1}{T}$ is the sample rate.

For the block of N samples

$$\underline{x}_2 = [x_2[1] \cdots x_2[N]]^T \quad (6.19)$$

we may write the vector equation

$$\underline{x}_2 = \alpha \underline{x}_1(s) + \underline{e}. \quad (6.20)$$

The reference channel data is represented by

$$\underline{x}_1(s) = \begin{bmatrix} x_1(1-s) \\ x_1(2-s) \\ \vdots \\ x_1(N-s) \end{bmatrix} \quad (6.21)$$

which is effectively a vector-function of time determined by band-limited interpolation of the observed reference channel samples $x_1[n]$.

If the error vector \underline{e} is zero-mean Gaussian and has auto-correlation matrix \mathbf{R}_{ee} then we may write the likelihood function

$$\mathcal{L}(\underline{x}_2; x_1(t), \alpha, s) = p(\underline{x}_2 | x_1(t), \alpha, s) \quad (6.22)$$

$$\propto \exp\left(-\frac{1}{2} \underline{e}^T \mathbf{R}_{ee}^{-1} \underline{e}\right) \quad (6.23)$$

$$\propto \exp\left(-\frac{1}{2} (\underline{x}_2 - \alpha \underline{x}_1(s))^T \mathbf{R}_{ee}^{-1} (\underline{x}_2 - \alpha \underline{x}_1(s))\right) \quad (6.24)$$

This formulation treats the first observed signal x_1 as a known vector-function of time. We model the random signal x_2 as being closely related to this function $x_1(t)$ by equation 6.17. Although we are dealing directly with the sampled representation $x_2[n]$ of $x_2(t)$, equation 6.22 still contains the continuous-time $x_1(t)$. We can evaluate this function for any value of t by band-limited interpolation of the signal samples $x_1[n]$.

6.5.1.1 Maximum Likelihood

The likelihood is not analytically differentiable, since $x_1(t)$ is a non-analytic function of t , and therefore calculation of the Maximum Likelihood offset estimate s_{ML} requires application of numerical maximisation techniques. The dimensionality of the problem is sufficiently small, however, that this is a realistic option. Furthermore, a numerical approximation to the differential can easily be calculated to assist in such an optimisation scheme.

6.6 Bayesian Formulation

The next step of sophistication is to treat the sequence of offsets measured by any of the preceding methods of sections 6.2 to 6.5 as a noisy observation

$$s_0[m] = s[m] + e_s[m] \quad (6.25)$$

of the true offset at frame \mathbf{m} , given by $s[m]$. The observation noise is represented by \underline{e}_s . The likelihood for the observed offset sequence is given by

$$\mathcal{L}(\underline{s}_0; \underline{s}) = p_{s_0|s, \mathcal{M}}(\underline{s}_0 | \underline{s}, \mathcal{M}) \quad (6.26)$$

$$\propto \exp\left(-\frac{1}{2} \underline{e}_s^T \mathbf{R}_{e_s}^{-1} \underline{e}_s\right) \quad (6.27)$$

$$\propto \exp\left(-\frac{1}{2} (\underline{s}_0 - \underline{s})^T \mathbf{R}_{e_s}^{-1} (\underline{s}_0 - \underline{s})\right) \quad (6.28)$$

assuming Gaussian errors with correlation matrix \mathbf{R}_{e_s} , and where the model structure is represented by \mathcal{M} .

Bayes' Rule allows us to combine this likelihood with a prior for the offset, giving the posterior p.d.f. of the offset \underline{s} given the observed data \underline{s}_0 as

$$p_{s|s_0, \mathcal{M}}(\underline{s} | \underline{s}_0, \mathcal{M}) = \frac{p_{s_0|s, \mathcal{M}}(\underline{s}_0 | \underline{s}, \mathcal{M}) p_{s, \mathcal{M}}(\underline{s}, \mathcal{M})}{p_{s_0, \mathcal{M}}(\underline{s}_0, \mathcal{M})} \quad (6.29)$$

from which we can make estimates of the true time offset.

We will assume the model structure \mathcal{M} from here on, and also that any parameters of that model are known. Furthermore, for a given set of observations $p_{s_0, \mathcal{M}}(\underline{s}_0, \mathcal{M})$ is constant, and we may therefore write

$$p_{s|s_0}(\underline{s} | \underline{s}_0) = \frac{p_{s_0|s}(\underline{s}_0 | \underline{s}) p_s(\underline{s})}{p_{s_0}(\underline{s}_0)} \quad (6.30)$$

$$\propto p_{s_0|s}(\underline{s}_0 | \underline{s}) p_s(\underline{s}) \quad (6.31)$$

where the prior p.d.f. $p_s(s)$ will depend upon the model that is chosen for the offset.

6.7 Models and Priors for the Offset

In the absence of further information a uniform prior may be chosen for $p_s(\underline{s})$, and in this case the posterior p.d.f. is equal to the likelihood. However, a consistency

of character has been noted in the offset variation (figures 6.2, 6.5) and this implies that there are common underlying mechanisms which can be exploited in a model-based approach.

We will consider now various specific forms for the model and this prior.

6.7.1 AR plus Sinusoids

The first model we propose for the sampled offset $s[nT_s]$ is the “AR plus sinusoidal basis” model which combines a stochastic component with an oscillatory component of fixed period³. This has been found to give highly satisfactory results in a variety of situations, and will therefore be examined in detail.

The offset shown in figure 6.5 seems to show a type of oscillatory behaviour, but with an additional random component such that it is not perfectly periodic, and which adds a long-term drift.

If we sample the offset at the rate $1/T_s$ the model equation is given by

$$s[n] = \sum_{k=1}^{P_a} a_k s[n-k] + \sum_{k=1}^{P_h} (\alpha_k \cos(k\omega_0 n T_s) + \beta_k \sin(k\omega_0 n T_s)) + e[n] \quad (6.32)$$

and this may be written for a set of N consecutive frames

$$\underline{s} = [s[1] \cdots s[N]]^T \quad (6.33)$$

as the matrix equation

$$\underline{s} = S\underline{a} + \Omega_c \underline{\alpha} + \Omega_s \underline{\beta} + \underline{e} \quad (6.34)$$

where

$$S = \begin{bmatrix} s[0] & \cdots & s[-P_a + 1] \\ s[1] & \cdots & s[-P_a + 2] \\ \vdots & & \vdots \\ s[N-1] & \cdots & s[-P_a + N] \end{bmatrix} \quad (6.35)$$

$$\Omega_c = \begin{bmatrix} \cos(\omega_0 T_s) & \cos(2\omega_0 T_s) & \cdots & \cos(P_h \omega_0 T_s) \\ \cos(\omega_0 2T_s) & \cos(2\omega_0 2T_s) & \cdots & \cos(P_h \omega_0 2T_s) \\ \vdots & \vdots & & \vdots \\ \cos(\omega_0 NT_s) & \cos(2\omega_0 NT_s) & \cdots & \cos(P_h \omega_0 NT_s) \end{bmatrix} \quad (6.36)$$

³From here on, n is used to denote the frame number, and T_s the frame period. The sample frames do not overlap.

$$\Omega_s = \begin{bmatrix} \sin(\omega_0 T_s) & \sin(2\omega_0 T_s) & \cdots & \sin(P_h \omega_0 T_s) \\ \sin(\omega_0 2T_s) & \sin(2\omega_0 2T_s) & \cdots & \sin(P_h \omega_0 2T_s) \\ \vdots & \vdots & & \vdots \\ \sin(\omega_0 NT_s) & \sin(2\omega_0 NT_s) & \cdots & \sin(P_h \omega_0 NT_s) \end{bmatrix} \quad (6.37)$$

and \underline{a} , $\underline{\alpha}$ and $\underline{\beta}$ are the vectors of weights for the AR, cosine and sine model components respectively.

6.7.1.1 AR+Sinusoidal Prior

We assume that, when the correct model parameters are known or have been accurately estimated, the error sequence \underline{e} is white and Gaussian. The Bayesian prior which corresponds to this signal model is therefore

$$p_s(\underline{s}) \propto \exp\left(-\frac{1}{2} \frac{\underline{e}^T \underline{e}}{\sigma_e^2}\right) \quad (6.38)$$

where \underline{e} is given as follows.

A trivial rearrangement of 6.32 gives

$$\mathbf{e}[\mathbf{n}] = \mathbf{s}[\mathbf{n}] - \sum_{k=1}^{P_a} \alpha_k \mathbf{s}[\mathbf{n} - k] - \sum_{k=1}^{P_h} (\mathbf{a}_k \cos(k\omega_0 \mathbf{n}T_s) + \mathbf{b}_k \sin(k\omega_0 \mathbf{n}T_s)) \quad (6.39)$$

and we may write this in matrix form

$$\underline{e} = \mathbf{A}\underline{s} - \Omega_c \underline{\alpha} - \Omega_s \underline{\beta} \quad (6.40)$$

where \mathbf{A} is the matrix of AR model coefficients defined in equations 2.35 and 2.36 on page 19.

Simple algebraic manipulation allows us to write this in a more compact form

$$\underline{e} = \begin{bmatrix} \mathbf{A} & -\Omega_c & -\Omega_s \end{bmatrix} \begin{bmatrix} \underline{s} \\ \underline{\alpha} \\ \underline{\beta} \end{bmatrix} \quad (6.41)$$

$$= \mathbf{M} \begin{bmatrix} \underline{s} \\ \underline{\alpha} \\ \underline{\beta} \end{bmatrix} \quad (6.42)$$

and hence the Gaussian prior as

$$p_s(\underline{s}) \propto \exp\left(-\frac{1}{2\sigma_e^2} \begin{bmatrix} \underline{s}^T & \underline{\alpha}^T & \underline{\beta}^T \end{bmatrix} \mathbf{M}^T \mathbf{M} \begin{bmatrix} \underline{s} \\ \underline{\alpha} \\ \underline{\beta} \end{bmatrix}\right). \quad (6.43)$$

6.7.2 Parameter Estimation

Least squares estimation of the parameters of 6.34 is straightforwardly linear by minimisation of $\underline{\mathbf{e}}^T \underline{\mathbf{e}}$ if ω_0 is known and is given by the solution of

$$\begin{bmatrix} \mathbf{S}^T \\ \Omega_c^T \\ \Omega_s^T \end{bmatrix} \begin{bmatrix} \mathbf{S} & \Omega_c & \Omega_s \end{bmatrix} \begin{bmatrix} \underline{\mathbf{a}} \\ \underline{\alpha} \\ \underline{\beta} \end{bmatrix} = \begin{bmatrix} \mathbf{S}^T \\ \Omega_c^T \\ \Omega_s^T \end{bmatrix} \underline{\mathbf{s}} \quad (6.44)$$

for the composite parameter vector $[\underline{\mathbf{a}}^T \ \underline{\alpha}^T \ \underline{\beta}^T]^T$.

However, it will be more usually the case that the fundamental frequency is not precisely known, and should therefore be treated as an unknown parameter, and the problem becomes non-linear in this parameter. Given a suitable starting value for ω_0 , iterative minimisation of the modelling error with respect to, alternately, the linear parameters (by equation 6.44) and the fundamental frequency (*e.g.* by Newton-Raphson iteration) quickly and reliably converges.

The prediction error is given by

$$\underline{\mathbf{e}} = \mathbf{S}\underline{\mathbf{a}} + \Omega_c \underline{\alpha} + \Omega_s \underline{\beta} - \underline{\mathbf{s}} \quad (6.45)$$

and its energy by $\mathcal{E} = \underline{\mathbf{e}}^T \underline{\mathbf{e}}$.

The differential of the energy with respect to ω , the fundamental frequency is given by

$$\frac{\partial \mathcal{E}}{\partial \omega} = 2 \left(\mathbf{S} \begin{bmatrix} \underline{\mathbf{a}} \\ \underline{\alpha} \\ \underline{\beta} \end{bmatrix} - \underline{\mathbf{s}} \right)^T \frac{\partial \Omega}{\partial \omega} \begin{bmatrix} \underline{\mathbf{a}} \\ \underline{\alpha} \\ \underline{\beta} \end{bmatrix} + \begin{bmatrix} \underline{\mathbf{a}} \\ \underline{\alpha} \\ \underline{\beta} \end{bmatrix}^T \left[\Omega^T \frac{\partial \Omega}{\partial \omega} + \frac{\partial \Omega^T}{\partial \omega} \Omega \right] \begin{bmatrix} \underline{\mathbf{a}} \\ \underline{\alpha} \\ \underline{\beta} \end{bmatrix} \quad (6.46)$$

where $\Omega = [\Omega_c \ \Omega_s]$. This is an analytic function and can therefore be used straightforwardly in the optimisation.

6.7.3 Test Results

Figure G.5 shows the prediction of the offset given by a simple model of this type. This example uses just one AR coefficient ($P_a = 1$) and two sinusoidal harmonics ($P_h = 2$). This model gives a lower prediction error than an AR model with an equivalent number of parameters ($P_a = 3, P_h = 0$), but the latter would require less computation.

The initial estimate of the period was 33.00 r.p.m. and this was refined to 33.34 r.p.m. at the end of the optimisation. Calculation of the model parameters in this example represents an overhead of 8×10^5 floating point operations over 30 seconds. This represents approximately 0.5% of the computational capacity of an inexpensive DSP chip, despite the non-linear iterative nature of the problem.

6.7.4 Alternative Offset Models and Priors

The AR plus sinusoid model has been shown to work well for the time offset that occurs as a result of turntable wow. There are other well-known mechanisms that can generate time offsets, such as the uneven stretch that affects magnetic tape that has been handled poorly. Furthermore there may be an overall speed discrepancy between the signals, or one may increase steadily in speed.

In these cases the AR plus sinusoid model may not perform well, and we suggest here some possible alternatives.

6.7.4.1 AR Model

The AR model provides a very general framework which, depending on its parameters, can model a wide range of offset characters. With poles close to the origin the modelled offset is highly random, whereas the model becomes quasi-periodic if the poles are close to the unit circle.

The formulation for the AR model is identical to the AR+sinusoid model with the number of sinusoidal harmonics P_h set to zero. In this case the Gaussian prior becomes

$$p_s(\underline{s}) \propto \exp\left(-\frac{1}{2}\underline{s}^T A^T A \underline{s}\right) \quad (6.47)$$

where A is the matrix of AR model coefficients defined in equations 2.35 and 2.36 on page 19.

6.7.4.2 Differential Smoothness Model

The offset may frequently be expected to vary smoothly, without sudden changes. In this case some objective measure of smoothness θ is maximised for the estimated offset variation. One such measure that has been suggested for the con-

tinuous case is the integral of derivatives

$$\theta = \int_0^{t_0} \left| \frac{d^q s(t)}{dt^q} \right|^2 dt \quad (6.48)$$

and in the discrete case this may be approximated in terms of the finite differences

$$d_n^1 = s_n - s_{n-1} \quad (6.49)$$

$$d_n^2 = d_n^1 - d_{n-1}^1 \quad (6.50)$$

and so on.

These differences may readily be expressed in matrix form and we can then form the sum of the squared order q differences as

$$\theta_d = \sum_{q+1}^N (d_n^q)^2 \quad (6.51)$$

$$= \underline{s}^T D_q^T D_q \underline{s} \quad (6.52)$$

where D_q is the matrix which generates order- q differences from the vector \underline{s} of length N .

Godsill has used this formulation in work on restoration of pitch defects on a single signal [38] and suggests the Gaussian prior

$$p(\underline{s}) \propto \exp\left(-\frac{\alpha \theta_d}{2}\right) \quad (6.53)$$

where α is set appropriately to reflect the expected degree of smoothness for a particular problem.

The differential smoothness model of order q may be shown to be a special case of the AR model with all of its poles at $z = 1$. It therefore removes the need for estimation of the model parameters.

6.7.4.3 Polynomial Model

Where there is a speed discrepancy between the two signals there will be a linear dependence $s(t) = \gamma t$ of the offset upon time. If this situation, or some other deterministic, non-oscillatory phenomenon is suspected then incorporation of such a term in the offset model and prior will be expected to yield substantially improved results.

More generally, polynomial terms of any degree may be incorporated by inclusion of the terms

$$s[n] = \sum_{i=0}^{P_p} \gamma_i \left(\frac{n}{T_s} \right)^i \quad (6.54)$$

in the offset predictor. This may be expressed in matrix form

$$\underline{s} = \mathbf{T}\boldsymbol{\gamma} \quad (6.55)$$

where

$$\mathbf{T} = \begin{bmatrix} 1 & \left(\frac{1}{T_s}\right) & \left(\frac{1}{T_s}\right)^2 & \cdots & \left(\frac{1}{T_s}\right)^{P_p} \\ 1 & \left(\frac{2}{T_s}\right) & \left(\frac{2}{T_s}\right)^2 & \cdots & \left(\frac{2}{T_s}\right)^{P_p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \left(\frac{N}{T_s}\right) & \left(\frac{N}{T_s}\right)^2 & \cdots & \left(\frac{N}{T_s}\right)^{P_p} \end{bmatrix}. \quad (6.56)$$

These deterministic terms may be readily incorporated in the same way as the sinusoidal terms in the AR+sinusoid model described in section 6.7.1, and the form of the prior is therefore identical.

6.8 Offset Posterior PDF

Combination of the offset measurements \underline{s}_0 and a prior based on a model \mathcal{M} using Bayes' rule allows determination of the posterior p.d.f. for the offset \underline{s} as

$$p_s(\underline{s} | \underline{s}_0, \mathcal{M}) = \frac{p_{s_0}(\underline{s}_0 | \underline{s}) p_s(\underline{s}, \mathcal{M})}{p_{s_0}(\underline{s}_0, \mathcal{M})} \quad (6.57)$$

Investigation of this p.d.f. allows *a-posteriori* estimates to be made of the offset.

6.8.1 MAP Offset Estimate

The posterior p.d.f. may be maximised for the MAP estimate

$$\underline{s}_{\text{MAP}} = \underset{\underline{s}}{\operatorname{argmax}} \{p_s(\underline{s} | \underline{s}_0)\} \quad (6.58)$$

of the offset variation.

Dependent on the choice of prior this maximisation may be analytic by standard differential calculus of vectors. For example, in the case of the AR+sinusoid prior

(equation 6.43) we obtain the posterior p.d.f.

$$p_s(\underline{s} \mid \underline{s}_0) \propto \exp \left(-\frac{(\underline{s}_0 - \underline{s})^\top (\underline{s}_0 - \underline{s})}{2\sigma_n^2} - \frac{1}{2\sigma_e^2} \begin{bmatrix} \underline{s}^\top & \underline{\alpha}^\top & \underline{\beta}^\top \end{bmatrix} \mathbf{M}^\top \mathbf{M} \begin{bmatrix} \underline{s} \\ \underline{\alpha} \\ \underline{\beta} \end{bmatrix} \right) \quad (6.59)$$

where \mathbf{M} is defined in equation 6.42.

This p.d.f. may be directly differentiated to give the offset estimate

$$\underline{s}_{MAP} = \left(\frac{\sigma_n^2}{\sigma_e^2} \mathbf{A}^\top \mathbf{A} + \mathbf{I} \right)^{-1} \left(\underline{s}_0 + \frac{\sigma_n^2}{\sigma_e^2} \mathbf{A}^\top (\Omega_c \underline{\alpha} + \Omega_s \underline{\beta}) \right) \quad (6.60)$$

6.8.2 Joint estimate of Model Parameters and Offset

It is possible to estimate the model parameters corresponding to fixed basis functions jointly with the offset in this framework. Let us define the vectors

$$\underline{\theta} = \begin{bmatrix} \underline{s} \\ \underline{\alpha} \\ \underline{\beta} \end{bmatrix} \quad (6.61)$$

$$\underline{\theta}_0 = \begin{bmatrix} \underline{s}_0 \\ \underline{\mu}_\alpha \\ \underline{\mu}_\beta \end{bmatrix} \quad (6.62)$$

and the matrix

$$\mathbf{R}_\theta^{-1} = \begin{bmatrix} \sigma_n^{-2} \mathbf{I} & 0 & 0 \\ 0 & \mathbf{R}_{\alpha\alpha}^{-1} & \mathbf{R}_{\alpha\beta}^{-1} \\ 0 & \mathbf{R}_{\beta\alpha}^{-1} & \mathbf{R}_{\beta\beta}^{-1} \end{bmatrix} \quad (6.63)$$

such that $\underline{\mu}_\alpha$, $\underline{\mu}_\beta$, $\mathbf{R}_{\alpha\alpha}$, $\mathbf{R}_{\alpha\beta}$, $\mathbf{R}_{\beta\alpha}$ and $\mathbf{R}_{\beta\beta}$ form the parameters of a Gaussian prior on $\underline{\alpha}$ and $\underline{\beta}$, the weights applied to the constant basis vectors.

The posterior p.d.f. for $\underline{\theta}$ is given by

$$p_s(\underline{\theta} \mid \underline{s}_0) \propto \exp \left(-\frac{1}{2} (\underline{\theta}_0 - \underline{\theta})^\top \mathbf{R}_\theta^{-1} (\underline{\theta}_0 - \underline{\theta}) - \frac{1}{2\sigma_e^2} \underline{\theta}^\top \mathbf{M}^\top \mathbf{M} \underline{\theta} \right) \quad (6.64)$$

from which the MAP estimate

$$\underline{\theta}_{MAP} = \begin{bmatrix} \underline{s}_{MAP} \\ \underline{\alpha}_{MAP} \\ \underline{\beta}_{MAP} \end{bmatrix} = \left(\mathbf{R}_\theta^{-1} + \frac{1}{\sigma_e^2} \mathbf{M}^\top \mathbf{M} \right)^{-1} \mathbf{R}_\theta^{-1} \underline{\theta}_0 \quad (6.65)$$

of the offset and model parameters is obtained by differentiation.

6.9 Tests of Model-Based Bayesian Estimator

The Bayesian estimator was tested using the raw offsets (shown in figure 6.5) measured by the model-enhanced correlation method described in section 6.4. This data was treated as the observation \underline{s}_0 , and a MAP offset estimate obtained using equation 6.31 with various priors.

6.9.1 AR+Sinusoidal Prior

The AR plus Sinusoidal basis prior gives the result shown in figure G.6. The curve in the figure is the result of the joint offset and parameter estimate given by equation 6.65.

	LS	MAP
α_1	0.5714	0.5686
α_2	0.1754	0.1737
β_1	0.2894	0.2886
β_2	0.8095	0.8055

TABLE 6.1: *Harmonic amplitudes; AR plus sinusoid offset model.*

The harmonic amplitudes estimated by the MAP procedure are compared with those obtained by the least squares algorithm (equation 6.44) in table 6.1. The two methods are seen to give harmonic amplitudes that are in close agreement.

The MAP offset curve appears, as expected, to be a noise-reduced version of the observed data. In particular, various points in the measured data which appear to be outliers have been effectively suppressed.

6.9.2 Differential Smoothness Prior

Figure G.7 shows the results using the differential smoothness prior. This is an attractive option since it does not require the estimation of any parameters. The example shown uses a second-order smoothness measure, and this has been found to be widely applicable. The single parameter α can be set by hand, and very intuitively relates to the degree of smoothness expected in the result. Curves for two values of α are shown, and the difference between them illustrates the effect of this parameter.

6.10 Audio Demonstration

The model-enhanced correlation algorithm was used to estimate the time offset throughout [9] and [10].

The starts of the two recordings were aligned as accurately as possible using the SADiE digital audio workstation [94]. This was accomplished by examining the waveform and aligning by eye one musical event that was clearly visible in the waveform near the start of the extract.

The model-enhanced algorithm was then used to measure the time offset between the two recordings using finite frames of samples from each as previously discussed. The measurement was made independently for each of the groove walls, and the two measured time offsets are shown in figure 6.5. The MAP offset was then estimated using the AR+sinusoidal basis prior.

Track [9] was then shifted to be in alignment with [10]. To demonstrate the effectiveness of the algorithm the sum of the unprocessed tracks is presented (track [11]) as well as the sum of the resynchronised signals [12].

The sum of the unprocessed signals shows high degrees of colouration due to time-varying comb-filtering. This artifact is not present in the sum of the resynchronised tracks.

6.11 Conclusions

We have examined a number of methods for estimating the offset between a pair of similar audio signals. Such a pair of signals might be transcriptions of two copies of the same gramophone disc, or two magnetic tapes that have been recorded on different machines.

A number of methods for estimating the offset for a single frame of data were investigated. It was shown that the offset may be measured to sub-sample accuracy by consideration of an oversampled discrete cross-correlation function.

The cross-correlation method for offset measurement was enhanced by pre-whitening the signals using an AR model framework. The cross-correlation of the excitation sequences was shown to give a more precise estimate of the offset than the cross-correlation of the signals themselves.

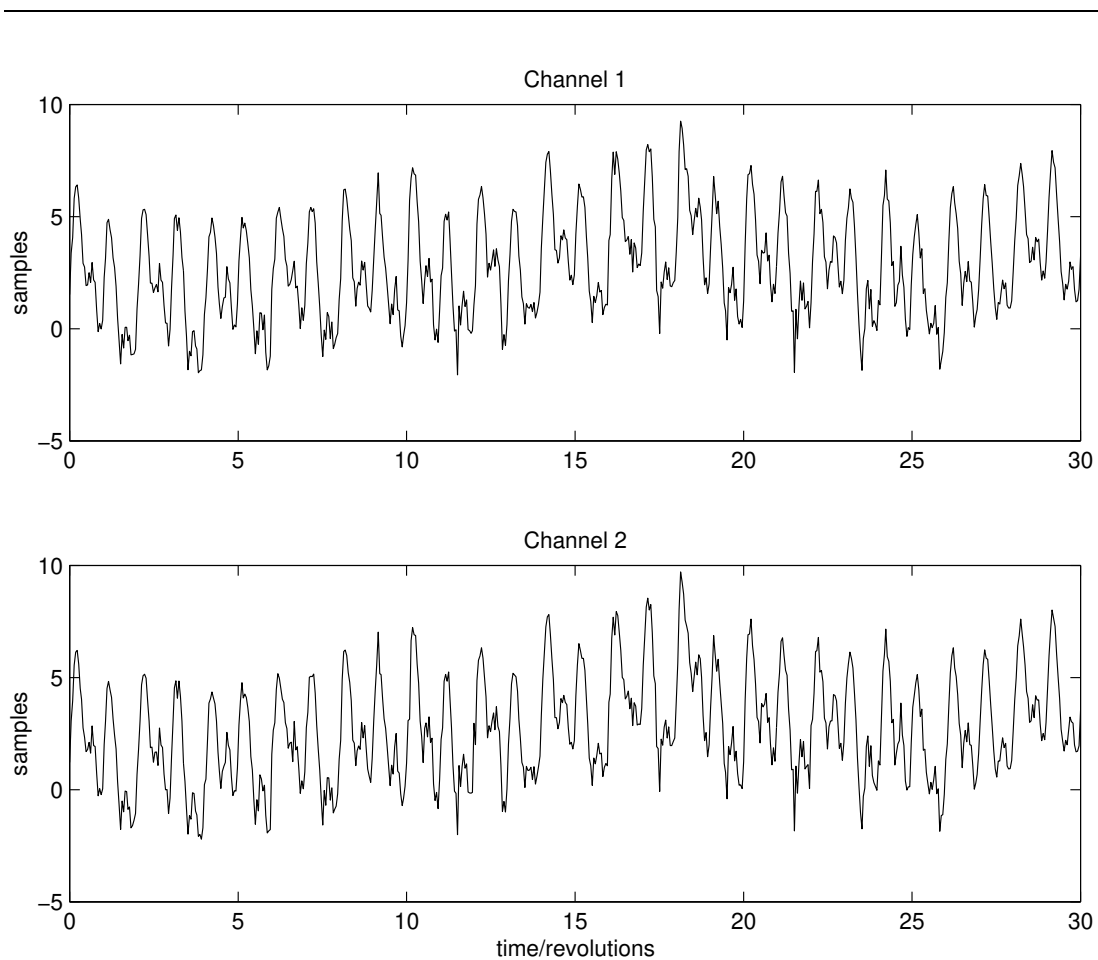


FIGURE 6.5: Time shift between two copies of pressing SPA-31. The time axis is calibrated in revolutions of a $33\frac{1}{3}$ r.p.m. gramophone record.

Model-based methods for regularising the measurements within a Bayesian framework were derived. These were shown to be robust for estimation of the offset over several frames. Various models were proposed for the offset variation, and it was shown how these can be incorporated as priors in the Bayesian estimation framework.

It is interesting to note that the turntable speed variations measured by Axon and Davies ([8], figure 5) in 1948, and those measured in the course of the present study are remarkably similar in character but that the modern turntable shows speed fluctuations an order of magnitude smaller (the speed difference between the transcriptions is obtained as the differential of the offset with respect to time). The former method is based on measurement of a constant tone on a single disc, while the present method is measuring the offset between two separate transcriptions.

Model-Based Quantisation

7.1	Scalar Quantisation	144
7.1.1	Dither	145
7.1.2	Noise-Shaped Quantisation	146
7.2	Model-Based Quantisation	150
7.3	Quantisation of Narrowband Signals	151
7.3.1	Dynamic Range Enhancement	151
7.3.2	Tonal Masking	154
7.3.3	Audio Demonstration	157
7.4	An Enhanced Linear Prediction Coder	157
7.4.1	Analysis of the LP CODEC	159
7.4.2	Application of Model-based Quantisation	160
7.4.3	Performance Tests	162
7.4.4	Coder Demonstration	164
7.5	Conclusions	164

Model-Based Quantisation

IN ORDER to manipulate signal samples using a computer each sample must be represented by a finite number of bits. Thus the resolution with which each sample is represented is not infinite. Amplitude errors are therefore introduced when a signal is converted to this digital form.

Quantisation occurs not just at the point of digitisation of the signal, but can also occur whenever the machine representation of a sample is changed. It will most often be significant when the number of bits used to represent a signal sample is reduced. This may be done as a deliberate part of, for example, a data compression algorithm. Alternatively it may be implicit such as when the 64-bit floating point result of a calculation is converted to a 16-bit integer for storage and transport on a compact disc.

7.1 Scalar Quantisation

Quantisation introduces errors into the signal, and the nature of these errors is dependent on the exact design of the quantiser and on the signal itself. We limit our consideration to quantisers of uniform step size, and we do not consider the effects of saturation of the digital word. We assume, unless stated otherwise and with no loss of generality, a stepsize of $q = 1$ throughout this chapter.

We may model the quantisation of a signal under these conditions as the process

of rounding a real number $x[n]$ to the nearest integer¹

$$x_q[n] = \left\lfloor x[n] + \frac{1}{2} \right\rfloor \quad (7.1)$$

where $\lfloor \cdot \rfloor$ represents the “floor” function which returns the greatest integer less than or equal to its argument².

The quantisation process may be abstracted further by modelling it as the addition of an error signal $e_q[n]$ such that

$$x_q[n] = x[n] + e_q[n]. \quad (7.2)$$

It can be shown that, under certain conditions, $e_q[n]$ has a uniform p.d.f. given by

$$p_{e_q}(e_q[n]) = \begin{cases} 1, & -\frac{1}{2} \leq e_q < \frac{1}{2}, \\ 0, & \text{otherwise} \end{cases} \quad (7.3)$$

The principal conditions for this to hold are that $x[n]$ is itself random, and that its amplitude p.d.f. spans a range much greater than the stepsize of the quantiser. Under these conditions the variance $\sigma_{e_q}^2 = \frac{1}{12}$.

7.1.1 Dither

In the basic quantiser the error $e_q[n] = x[n] - x_q[n]$, although random, may be highly correlated with the signal samples $x[n]$ [43, 87]. If x is itself a highly correlated audio signal then this is particularly undesirable as the effect of the quantisation is more akin to audible distortion of the signal than the addition of noise [68]. A demonstration of this phenomenon is given in [46].

This problem is readily circumvented by the addition of a random *dither* signal d prior to quantisation [87] as shown in figure 7.1 such that

$$x_q[n] = \left\lfloor x[n] + d[n] + \frac{1}{2} \right\rfloor. \quad (7.4)$$

¹Note that it is the default behaviour of many programming languages, including C, to round floating point numbers towards zero when performing an implicit floating point to integer conversion.

²When equation 7.1 is implemented as a computer algorithm there will be a small bias introduced as a result of the representation of the real number $x[n]$ in a finite floating-point format. In typical audio applications this bias is negligible.

In audio applications a white signal with the triangular p.d.f.

$$p_d(d[n]) = \begin{cases} 1 + d, & -1 \leq d < 0, \\ 1 - d, & 0 \leq d < 1, \\ 0 & \text{otherwise} \end{cases} \quad (7.5)$$

is frequently chosen for the dither signal (for example see [66, 31, 105]). This gives improved audio quality by ensuring that the first moment $E[e_q[n]] = 0$ and the second moment $\sigma_{e_q}^2 = E[e_q[n]^2]$ are independent of the original signal samples $x[n]$ [66].

Since $E[e_q[n]]$ is zero, independent of $x[n]$, the quantiser can be said to have been linearised by the dither, since

$$E[x_q[n]] = E[x[n] + e_q[n]] \quad (7.6)$$

$$= x[n] + E[e_q[n]] \quad (7.7)$$

$$= x[n] \quad (7.8)$$

The penalty for dithering the quantiser in this manner is that the noise power of the quantised signal is increased. The error variance for a quantiser incorporating this TPDF dither is $\sigma_{e_q}^2 = \frac{1}{4}$, and is 4.77 dB greater than for the undithered quantiser [66].

7.1.2 Noise-Shaped Quantisation

Consider the system shown in figure 7.2, in which we have added a feedback loop containing a filter with transfer function $H(z)$ around the dithered quantiser (comprising the additive dither source and the quantiser Q). Systems similar to this have been used extensively for analogue to digital and digital to analogue

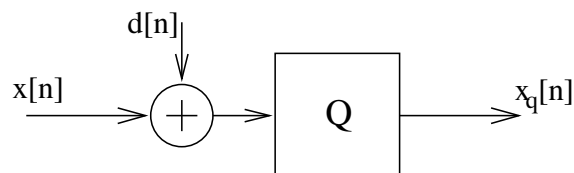


FIGURE 7.1: *Dithered Quantiser*

converters (for example [45, 95]), signal coding applications (for example [58]) and more recently baseband digital audio systems (for example [31, 103, 32, 104]).

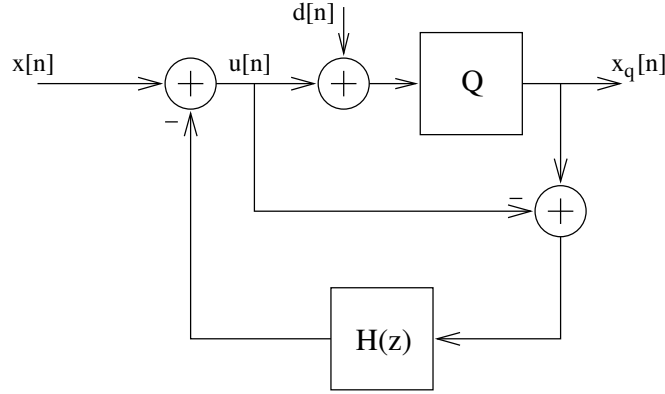


FIGURE 7.2: *Noise-Shaped Quantiser*

The output from the system is quantised (we continue to assume a stepsize of 1) and may be derived as follows. From the block diagram it is clear that

$$\mathbf{u}[\mathbf{n}] = \mathbf{x}[\mathbf{n}] - (\mathbf{x}_q[\mathbf{n}] - \mathbf{u}[\mathbf{n}]) \star \mathbf{h}[\mathbf{i}] \quad (7.9)$$

where \star represents the discrete convolution operator, and $\mathbf{h}[\mathbf{i}]$ is the impulse response

$$\mathbf{h}[\mathbf{i}] \stackrel{z}{\Leftrightarrow} \mathbf{H}(z) \quad (7.10)$$

of the feedback filter.

It is also clear that

$$\mathbf{x}_q[\mathbf{n}] = \mathbf{u}[\mathbf{n}] + \mathbf{e}_q[\mathbf{n}] \quad (7.11)$$

where the dithered quantisation function is modelled by the addition of $\mathbf{e}_q[\mathbf{n}]$ as before.

A trivial re-arrangement gives

$$\mathbf{u}[\mathbf{n}] = \mathbf{x}_q[\mathbf{n}] - \mathbf{e}_q[\mathbf{n}] \quad (7.12)$$

and substitution for $\mathbf{u}[\mathbf{n}]$ in equation 7.9 gives

$$\mathbf{x}_q[\mathbf{n}] - \mathbf{e}_q[\mathbf{n}] = \mathbf{x}[\mathbf{n}] - \mathbf{e}_q[\mathbf{n}] \star \mathbf{h}[\mathbf{i}] \quad (7.13)$$

$$= \mathbf{x}[\mathbf{n}] - \sum_{\mathbf{i}=-\infty}^{+\infty} \mathbf{h}[\mathbf{i}] \mathbf{e}_q[\mathbf{n} - \mathbf{i}]. \quad (7.14)$$

For the system to be realisable we require that

$$h[i] = 0, \quad i \leq 0 \quad (7.15)$$

and hence

$$x_q[n] - e_q[n] = x[n] - \sum_{i=1}^{+\infty} h[i] e_q[n - i]. \quad (7.16)$$

We may re-arrange this further to give the system output signal

$$x_q[n] = x[n] + e_q[n] - \sum_{i=1}^{\infty} h_i e_q[n - i] \quad (7.17)$$

$$= x[n] + \sum_{i=0}^{\infty} h'_i e_q[n - i] \quad (7.18)$$

where

$$h'[i] = \begin{cases} 0, & i < 0 \\ 1, & i = 0 \\ -h_i, & i > 0 \end{cases} \quad (7.19)$$

The filter h' has the z -domain transfer function

$$h'[i] \stackrel{z}{\Leftrightarrow} 1 - H(z) \quad (7.20)$$

where

$$h[i] \stackrel{z}{\Leftrightarrow} H(z) \quad (7.21)$$

is the actual filter implemented in the system.

The quantisation error that appears in the output signal (equation 7.18) is filtered by the function $(1 - H(z))$. Hence the power spectral density of the noise component of the output signal is no longer white, but has been shaped by the noise-shaping function $(1 - H(z))$ such that

$$S_{ee}(\omega) = |1 - H(e^{-j\omega T})|^2 \sigma_{e_q}^2. \quad (7.22)$$

The filter $H(z)$ is frequently (but not necessarily) chosen to be the FIR filter

$$H(z) = \sum_{i=1}^{P_h} h[i] z^{-i}. \quad (7.23)$$

The summation starts from $i = 1$ for realisability as discussed above. In this case the effective noise-shaping function becomes

$$1 - H(z) = \sum_{i=0}^{P_h} h'_i z^{-i}. \quad (7.24)$$

Note that this choice of an FIR filter is merely a convenience, and not a restriction, and in principle noise-shaping systems with recursive feedback filters are entirely practical.

There are a number of points to note regarding this noise-shaped quantiser:

- There is an implicit term $h'_0 = 1$ in the noise shaping function. As a result, the noise shaping function $(1 - H(z))$ cannot be chosen completely arbitrarily.
- If $H(z)$ itself is FIR (and is therefore unconditionally stable) then so is the complete system unconditionally stable, despite having a feedback path. This is clear from consideration of the transfer functions from each of $X(z)$ and $E(z)$ to $X_q(z)$, neither of which contains poles.
- The system has a recursive nature and can therefore suffer from limit cycles and idle tones when implemented in finite-precision arithmetic. This statement holds even if $H(z)$ itself is FIR. The dither helps to alleviate this problem [24, 67].

None of these points represents a significant hindrance to the implementation of such a noise shaper, and indeed this topology is used widely in many audio applications.

The total noise power \mathcal{P} at the output of the noise-shaped quantiser is given by

$$\mathcal{P} = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \sigma_{e_q}^2 |1 - H(e^{-j\theta})|^2 d\theta. \quad (7.25)$$

It was shown above that for a system using triangular p.d.f. dither $\sigma_{e_q}^2 = \frac{1}{4}$. Furthermore, if an FIR filter is chosen for $H(z)$ the noise power gain may be readily evaluated as the sum of the squares of the filter coefficients [46], giving the noise output power

$$\mathcal{P} = \frac{1}{4} \sum_{i=0}^P h'_i{}^2 \quad (7.26)$$

where h' is as defined before.

Equation 7.26 implies a law of diminishing returns. As we try to exercise increasing control on the noise power spectrum by increasing the length P of the noise-shaping filter, so the total noise power rises.

Note that $1 - H(z)$ should ideally be chosen to be a minimum-phase design, as this gives minimum output noise power for a given noise spectral shape. For example, a pair of conjugate minimum-phase zeros $z = re^{\pm j\theta}$ ($r^2 < 1$) is given by

$$1 - H_1(z) = 1 - z^{-1}2r \cos \theta + z^{-2}r^2 \quad (7.27)$$

and the same amplitude response is given by the non-minimum-phase filter

$$1 - H_2(z) = 1 - z^{-1}2\frac{1}{r} \cos \theta + z^{-2} \left(\frac{1}{r}\right)^2. \quad (7.28)$$

For $r^2 < 1$ it is clear that equation 7.26 will evaluate to a greater noise power total for the non-minimum-phase filter (equation 7.28) than for the minimum-phase filter (equation 7.27).

This does not necessarily imply that $H(z)$ itself will be a minimum-phase filter. In our second-order example the zero of $H_1(z)$ is at $z = \frac{r}{2 \cos \theta}$, which is not constrained to be within the unit circle.

7.2 Model-Based Quantisation

We can tailor the noise-shaping function adaptively in accordance with the signal we are quantising. This makes the noise-shaping filter $H(z)$ some (non-trivial) function of the signal samples $x[n]$, and hence can alter the shape of the quantisation noise spectrum depending on the signal content.

If we have a model for the signal then we may base the filter upon the parameters of that model. The arrangement is shown diagrammatically in figure 7.3. We term this new extension to the noise-shaped quantiser “model-based quantisation”.

Possibilities include the use of AR or ARMA model filters as the feedback filter $H(z)$. Either the forward or the inverse forms of the filters may be used, provided that the system remains causal and stable. In an adaptive system it is particularly attractive to use the FIR form for $H(z)$ since, as noted above, it guarantees that the system will be stable. It is also important to remember that it is desirable

that the noise-shaping function (and not the filter itself) be minimum-phase, as this minimises the noise amplification of the system.

The model-based quantiser is not inherently limited in application to audio signals, but for illustration two audio applications based on the AR model are shown in the following sections.

7.3 Quantisation of Narrowband Signals

Many audio signals are relatively narrow-band compared with the Nyquist bandwidth for the sample rate being used. The first application of the model-based quantiser is to manipulate the quantisation noise adaptively according to the signal characteristics. In particular we show how the noise may be either reduced in those areas of the passband which are occupied by the signal, or alternatively concentrated close to the strong signal components.

We may wish to do either of these things in different circumstances.

- If we have a signal which we are transmitting over a channel and we wish to recover it at the far end using some form of adaptive filter, then the recovered signal will be improved if we can reduce the noise power in the passband of the filter. This can be accomplished by pushing the noise away from the signal components using the model-based quantiser.
- In quantising an audio signal we may wish to exploit the phenomenon of tonal masking [73] to hide the quantisation noise. Perhaps this may be accomplished by using the model-based quantiser to move the noise close to the tonal components of the signal.

If the signal may be modelled as auto-regressive then the AR parameters form the basis of filters $H(z)$ which, incorporated in the model-based quantiser, will perform either of these functions.

7.3.1 Dynamic Range Enhancement

Let us suppose that we have an AR signal

$$x[n] = e[n] + \sum_{i=1}^{P_a} a_i x[n-i] \quad (7.29)$$

and that we wish to quantise the samples $x[n]$. If we use the noise-shaped quantiser of figure 7.2 then the quantised signal is given *via* equation 7.18 as

$$x[n] = e[n] + \sum_{i=1}^{P_a} a_i x[n-i] + \sum_{i=0}^{P_h} h'_i e_q[n-i] \quad (7.30)$$

where addition of $e_q[n]$ represents the noise added by the dithered quantiser, and $h'_i \stackrel{z}{=} 1 - H(z)$ is the noise-shaper impulse response as previously defined.

It was shown above that the power spectral density (p.s.d.) at the system output comprises the signal, given by

$$S_{xx}(\omega) = \sigma_e^2 \left| \frac{1}{1 - \sum_{i=1}^{P_a} a_i e^{-j\omega iT}} \right|^2 \quad (7.31)$$

where P_a is the AR model order, and the noise

$$S_{nn}(\omega) = \frac{1}{4} \left| 1 - \sum_{i=1}^{P_h} h_i e^{-j\omega iT} \right|^2, \quad (7.32)$$

which is dependent on the noise-shaping filter $H(z)$.

If we set $H(z) = A(z)$ (and hence $P_h = P_a$) then the expression for the noise p.s.d. becomes

$$S_{nn}(\omega) = \frac{1}{4} \left| 1 - \sum_{i=1}^{P_a} a_i e^{-j\omega iT} \right|^2 \quad (7.33)$$

and by substitution from equation 7.31 we obtain

$$S_{nn}(\omega) = \frac{1}{4} \frac{\sigma_e^2}{S_{xx}(\omega)}. \quad (7.34)$$

We have succeeded in moving the quantisation noise away from parts of the spectrum occupied by the signal since there is now a reciprocal relationship between the signal p.s.d. $S_{xx}(\omega)$ and the quantisation noise p.s.d. $S_{nn}(\omega)$.

The relationship between the signal and noise spectra is illustrated in figure 7.4. The upper graph shows the power spectrum of a two-tone test signal quantised using a straightforward dithered quantiser. The noise floor is flat, and shows no distortion spurious. The lower graph shows the output of the model-based quantiser, which is set to the same stepsize. Notice that the quantisation noise has been moved away from the area of the spectrum occupied by the signal.

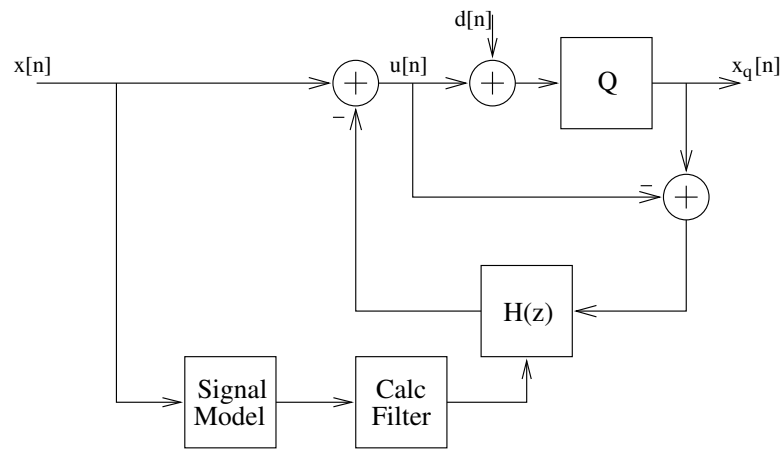


FIGURE 7.3: Model-based Quantiser

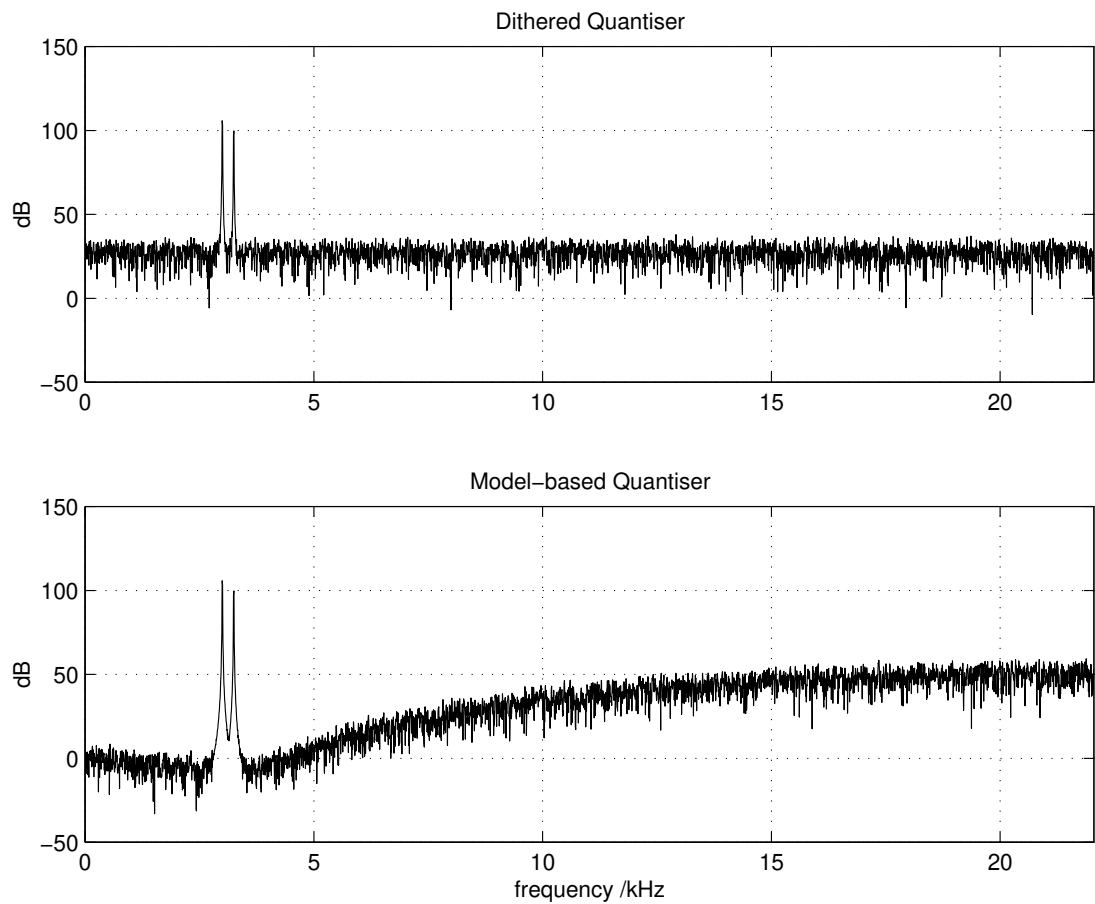


FIGURE 7.4: Model-based Quantisation of a Narrowband Signal

The total noise power \mathcal{P} in the resultant signal may be calculated from equation 7.26 and is given by

$$\mathcal{P} = \frac{1}{4} \left(1 + \sum_{i=1}^p a_i^2 \right). \quad (7.35)$$

Assuming the model parameters $\underline{\mathbf{a}} = [a_1 \cdots a_p]^T$ are estimated by the correlation method described in section 2.4 then this noise power may be expressed in terms of the signal autocorrelation

$$\mathcal{P} = \frac{1}{4} (1 + \underline{\mathbf{a}}^T \underline{\mathbf{a}}) \quad (7.36)$$

$$= \frac{1}{4} (1 + \underline{\mathbf{r}}^T \mathbf{R}^{-T} \mathbf{R}^{-1} \underline{\mathbf{r}}) \quad (7.37)$$

where \mathbf{R} and $\underline{\mathbf{r}}$ are defined in section 2.4.2, and \mathbf{R}^{-T} denotes the inverse transpose of matrix \mathbf{R} .

It was noted in section 2.4 that the autocorrelation method is guaranteed to generate a stable AR model. This is equivalent to the property that the noise-shaping function in the model-based quantiser be minimum phase, since poles in the AR model are converted directly to zeros in the noise shaping function.

This was shown in section 7.1.2 to be a desirable feature, since it guarantees minimum output noise power for a given noise floor shape. Thus this model-based quantiser is guaranteed to give the optimum noise-shaping filter for the noise p.s.d. it generates.

7.3.2 Tonal Masking

In order to make the quantisation noise appear close to the tonal components of the signal we can arrange that the noise has the same spectral shape as the AR spectrum of the signal. In order to achieve this we require that the noise-shaping function is equal to the transfer function of the model filter

$$1 - H(z) = \frac{1}{1 - A(z)}. \quad (7.38)$$

Some manipulation determines that the filter $H(z)$ is then given by

$$H(z) = \frac{-A(z)}{1 - A(z)}. \quad (7.39)$$

Note that although this filter is not FIR, it is stable (if the model from which it is derived is stable), and that the noise-shaping function $1 - H(z)$ is minimum-phase,

as was noted above to be a desirable feature. The close relationship between the numerator and denominator polynomials gives an efficient filter structure which requires just $P + 2$ MAC³ operations.

Incorporation of this filter in the quantiser yields results shown in figure 7.5. The upper part of the figure shows the spectrum of a two-tone signal which has been quantised using a conventional dithered quantiser. The lower part shows the spectrum at the output of the model-based quantiser, and it is clear that the quantisation noise has been concentrated near the signal components, and reduced elsewhere.

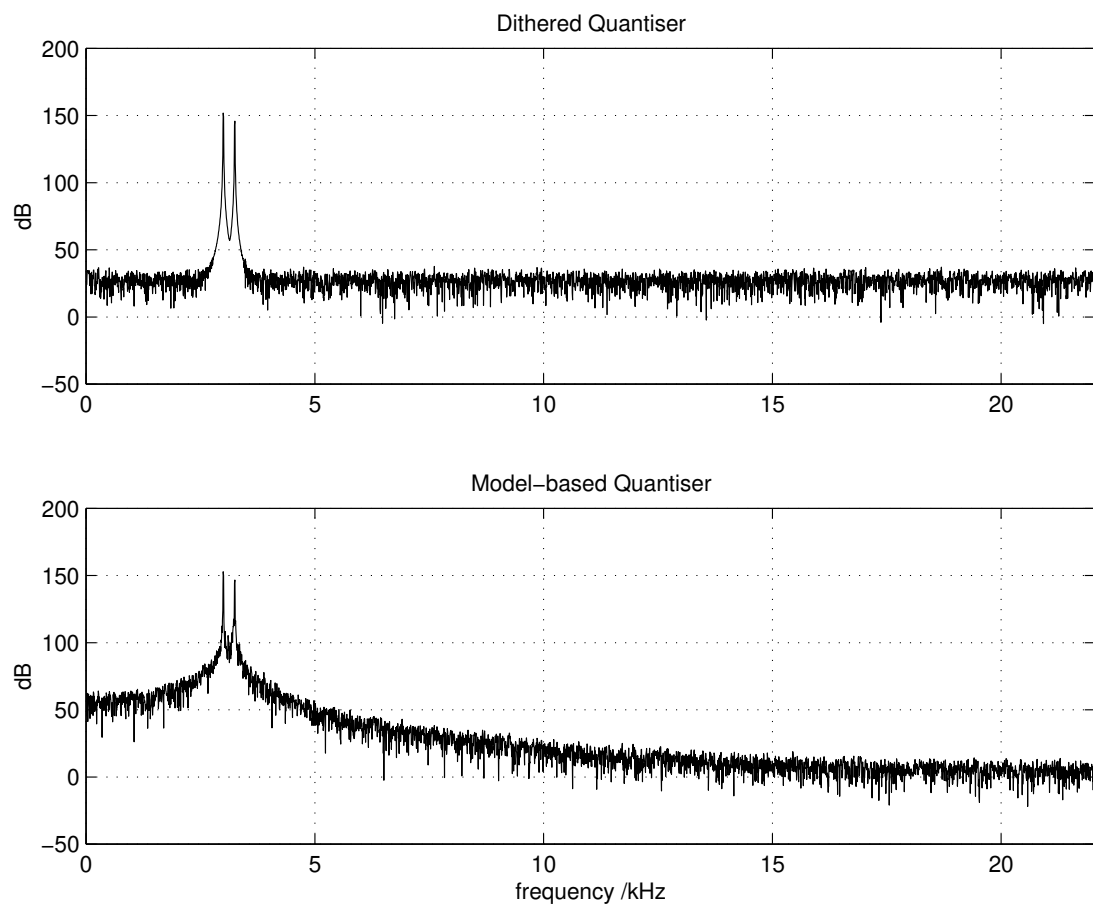


FIGURE 7.5: *Model-based Quantisation of a Narrowband Signal*

³One Multiply-Accumulate operation calculates $a \leftarrow a + b \cdot c$. This function is typically provided as a single DSP instruction.

The output noise p.s.d. is given by

$$S_{nn}(\omega) = \frac{1}{4} \left| \frac{1}{1 - \sum_{i=1}^{P_a} a_i e^{-j\omega i T}} \right|^2 \quad (7.40)$$

which may be compared with the signal p.s.d.

$$S_{xx}(\omega) = \sigma_e^2 \left| \frac{1}{1 - \sum_{i=1}^{P_a} a_i e^{-j\omega i T}} \right|^2. \quad (7.41)$$

The signal to noise ratio SNR_M at the output of the model-based quantiser is therefore given by

$$\text{SNR}_M = \frac{S_{xx}}{S_{nn}} \quad (7.42)$$

$$= 4\sigma_e^2 \quad (7.43)$$

which is independent of the signal spectrum shape. Note that the model prediction error energy σ_e^2 appears in this expression.

In some situations the effect of auditory masking results in the model-quantised signal sounding “cleaner” than the other, despite having a poorer signal to noise ratio. It has been found in many cases, however, that the increased noise power undoes the beneficial effect of the tonal masking. The noise power at lower frequencies than the signal peak is thought to be more audible than the noise slightly higher in frequency, since the masking effect of a tonal signal is greater at frequencies higher than the tone than at frequencies lower than the tone [73].

Furthermore, the output SNR is degraded as σ_e^2 becomes smaller (that is, as the model becomes a better description of the signal). This in some ways implies that this is a poor algorithm, since as the model becomes a better predictor of the data, so the auditory performance of the algorithm is degraded.

A possible counter these problems would be to artificially degrade the resonances of the AR model. By reducing the gain and Q of the model resonances the noise-shaping effect will be reduced in magnitude but without altering the centre frequencies of the peaks. The obvious method to accomplish this end is to solve the pole polynomial $1 - A(z) = 0$, and move each of the conjugate pole-pairs away from the unit circle. This is, however, computationally expensive for non-trivial model orders.

These issues do not dictate that the model-based quantiser has no application in this area, but do suggest that the particular filter we have examined, while sometimes useful, is not generally suitable.

7.3.3 Audio Demonstration

The accompanying CD contains an audio demonstration of these applications, but it should be noted that the technique is not inherently applicable only to audio signals.

Tracks [13]–[15] demonstrate the quantisation of a narrowband signal using the model-based quantiser. Track [13] serves as a reference, and consists of two tones. The first is at a constant 1 kHz, while the second rises in steps from approximately 100 Hz to a little over 3 kHz. Both decay gradually in amplitude.

Track [14] is this same signal quantised to an effective resolution of 8 bits using the conventional dithered quantiser of figure 7.1. The signal to noise ratio of the resulting tones decays from approximately 35 dB to -15 dB as they decay. Notice that the noise floor remains absolutely constant in perceived colour and level. The tones become almost inaudible at the end as they are masked by the quantisation noise.

Track [15] is the two-tone signal quantised to the same resolution using the model-based quantiser of figure 7.3 with $H(z) = A(z)$. At the start the low-frequency noise is audibly attenuated, compared with track [14], and the colour changes as the frequencies of the tones vary. Notice also that the tones are more easily perceptible at the end of track [15] than at the end of [14] due to the shaping of the noise floor away from the tonal components of the signal.

Track [16] demonstrates the concentration of noise near the signal components. The quantisation noise is audibly highly coloured, and its spectrum varies with the signal spectrum. It is highly audible due to the fact that the noise shaping filter is overly aggressive as discussed above.

7.4 An Enhanced Linear Prediction Coder

Linear Prediction (LP) Coding is based on the fact that for many signals, such as speech and music, each signal sample may be predicted, with some degree of accuracy, as the linear sum of P previous signal samples.

The prediction weights are calculated adaptively, on a block-by-block basis, in the coder so as to minimise the prediction error energy over the current finite block of samples. The prediction error samples are quantised to a shorter wordlength than was the original signal. These quantised error samples plus the calculated

weights for the block are then transmitted to the decoder.

The decoder reconstructs the original signal by filtering the quantised error samples with a reconstruction filter formed from the LP coefficients. The system is shown as a block diagram in figure 7.6.

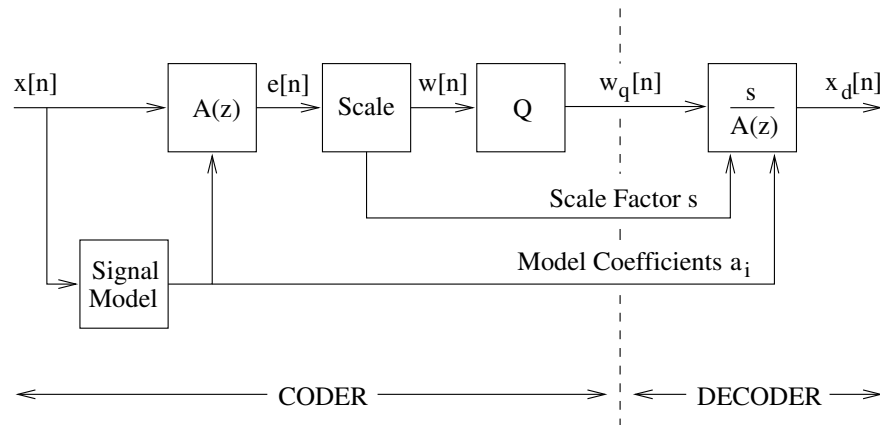


FIGURE 7.6: *Linear Prediction CODEC system*

Quantisation of the error sequence introduces quantisation noise and results in imperfect reconstruction of the coded signal. The enhancement to the basic LP coder presented here uses the model-based quantiser to whiten and reduce the amplitude of the errors in the reconstructed signal from this simple CODEC⁴ system.

We ignore for the present discussion two factors which affect the overall system performance:

- The linear prediction coefficients will be calculated in the coder in (typically) a 32 or 64 bit floating point format. In a specific CODEC implementation, however, they may be transmitted at a lower precision. Since the overhead of transmitting these coefficients is small in the system we describe we ignore the effect of this quantisation of the coefficients.
- The action of the model-based quantiser slightly increases the amplitude of the error signal w . Therefore the scale factor s will have to be smaller for the system using the model-based quantiser than for the basic system. The noise floor of the decoded signal is raised slightly, but this effect is very small (typically less than 1 dB).

⁴The combination of a “COder” and “DECOder” is frequently referred to as a “CODEC”.

The system analysis is largely unaffected by these simplifications, but an implementation of this system would, of course, require that attention be paid to both of these details.

7.4.1 Analysis of the LP CODEC

Suppose the signal we wish to code may be modelled as the auto-regressive signal

$$x[n] = s w[n] + \sum_{i=1}^{P_a} a_i x[n-i] \quad (7.44)$$

where $w[n]$ is a white excitation sequence. The factor s is chosen for a given block of data such that $w[n]$ is scaled suitably for the quantiser, as shown in figure 7.6.

From the definition of the AR model the optimal LP coefficients which minimise the prediction error are the model coefficients a_i , and thus the prediction error signal is identical to the excitation sequence $e[n] = s w[n]$ for this AR signal.

Quantisation of the white error signal may be modelled by the addition of uncorrelated white noise $e_q[n]$, of variance $\sigma_{e_q}^2 = \frac{1}{12}$. Note that we choose not to use dither in this application; the signal being quantised is approximately white, and so the quantisation will itself add white noise without the use of dither. Thus the transmitted error sequence

$$w_q[n] = w[n] + e_q[n] \quad (7.45)$$

is the sum of the scaled white excitation signal w and the white quantisation noise e_q .

At the decoder the quantised error sequence is applied to the reconstruction filter. The decoded signal is therefore given by

$$x_d[n] = s w_q[n] + \sum_{i=1}^{P_a} a_i x_d[n-i] \quad (7.46)$$

$$= s (e_q[n] + w[n]) + \sum_{i=1}^{P_a} a_i x_d[n-i] \quad (7.47)$$

$$= x[n] + s e_q[n] + \sum_{i=1}^{P_a} a_i x_d[n-i] \quad (7.48)$$

and the error in the decoded signal is

$$x_d[n] - x[n] = s e_q[n] + \sum_{i=1}^{P_a} a_i x_d[n-i]. \quad (7.49)$$

Notice that the white quantisation noise e_q is filtered by the reconstruction filter such that the resulting reconstruction error $x_d - x$ has the same power spectral shape as the original signal x .

The noise power \mathcal{P}_Q of the decoded signal is given by

$$\mathcal{P}_Q = \frac{\sigma_{e_q}^2 s^2}{2\pi} \int_{-\pi}^{\pi} \left| \frac{1}{1 + \sum_{i=1}^P a_i e^{-j\theta}} \right|^2 d\theta \quad (7.50)$$

where $\sigma_{e_q}^2 = \frac{1}{12}$, and the signal to noise ratio at the decoder output is given by

$$\text{SNR}_Q = \frac{\sigma_x^2}{\mathcal{P}_Q}. \quad (7.51)$$

7.4.2 Application of Model-based Quantisation

Many schemes have been put forward to improve the performance of this basic CODEC ([58]). The new enhancement presented here replaces the quantiser in the LP coder with the model-based quantiser of figure 7.3. Once again we choose the filter $H(z) = A(z)$ for the quantiser. Note that we have made no change to the decoder, as may be seen from figure 7.6.

The error signal transmitted by the coder, which now incorporates the model-based quantiser, is given by

$$w_q[n] = w[n] + \sum_{i=0}^{P_a} a'_i e_q[n - i] \quad (7.52)$$

where $a'_0 = 1$ and $a'_i = -a_i$, $1 \leq i \leq P_a$.

By comparison with equation 7.45 it can be seen that the quantisation error component of the transmitted error sequence w_q is no longer white, but rather has been shaped by a function of the signal model. Note that incorporation of the noise-shaper into the coder is efficient, requiring approximately $P + 1$ additional MAC operations per sample to apply the filter itself; the filter coefficients have already been calculated as a part of the original coder.

The decoder is unchanged, and thus the decoded signal, as before, is obtained by applying the quantised excitation to the reconstruction filter formed from the LP coefficients, and is therefore given by

$$x_d[n] = s w_q[n] + \sum_{i=1}^{P_a} a_i x_d[n - i] \quad (7.53)$$

The output signal may most easily be analysed by transforming into the z -domain. It may, at first, seem incorrect to do so, owing to the stochastic nature of the signals; however, since we are processing a finite, known block of data, we may treat this observed data as a set of known constants whose z -transforms are well-defined.

Taking the z -transforms of equations 7.52 and 7.53 we obtain

$$W_q(z) = W(z) + E_q(z)(1 - A(z)) \quad (7.54)$$

and

$$X_d(z) = s \frac{W_q(z)}{1 - A(z)} \quad (7.55)$$

respectively. Eliminating $W_q(z)$ gives

$$X_d(z) = s \frac{W(z) + E_q(z)(1 - A(z))}{1 - A(z)} \quad (7.56)$$

$$= X(z) + s E_q(z) \frac{1 - A(z)}{1 - A(z)} \quad (7.57)$$

$$= X(z) + s E_q(z) \quad (7.58)$$

and hence *via* the inverse z -transform

$$x_d[n] = x[n] + s e_q[n]. \quad (7.59)$$

This output signal comprises the desired original signal $x[n]$, and the *white* interference signal $s e_q[n]$.

The output noise power and signal to noise ratio are given by

$$\mathcal{P}_M = \sigma_{e_q}^2 s^2 \quad (7.60)$$

$$\text{SNR}_M = \frac{\sigma_x^2}{\sigma_{e_q}^2 s^2}, \quad (7.61)$$

where once again $\sigma_{e_q}^2 = \frac{1}{12}$. The system therefore represents an improvement of

$$\frac{\text{SNR}_M}{\text{SNR}_Q} = \frac{1}{2\pi} \left(\int_{-\pi}^{\pi} \left| \frac{1}{1 + \sum_{i=1}^P \alpha_i e^{-j\theta}} \right|^2 d\theta \right)^{-1} \quad (7.62)$$

over the basic CODEC.

The model-based quantiser can never give worse performance than the basic quantiser in this application. In the worst case all the predictor coefficients α_i collapse to zero, and equation 7.62 equates to unity. Once again we note that the noise-shaping filter is guaranteed to be minimum-phase and hence is the optimum for its associated noise shape.

7.4.3 Performance Tests

The performance improvement clearly depends on the success with which the model predicts the data. For resonant audio signals such as speech or music, the error signal is typically reduced by 30–60 dB and very effectively whitened.

The enhanced CODEC was compared with the simple CODEC for the synthetic AR signal shown as the upper graph in figure 7.7. This was compressed by 4:1 in each case, and the decoded signals are shown as the lower part of the same figure.

Both CODECs seem to have preserved the nature of the signal, but the performance difference becomes very much clearer by examination of the reconstruction error, shown in figure 7.8. The upper graph shows the error resulting from the basic CODEC, while the lower shows, on the same scale, the error when the model-based quantiser is substituted in the coder.

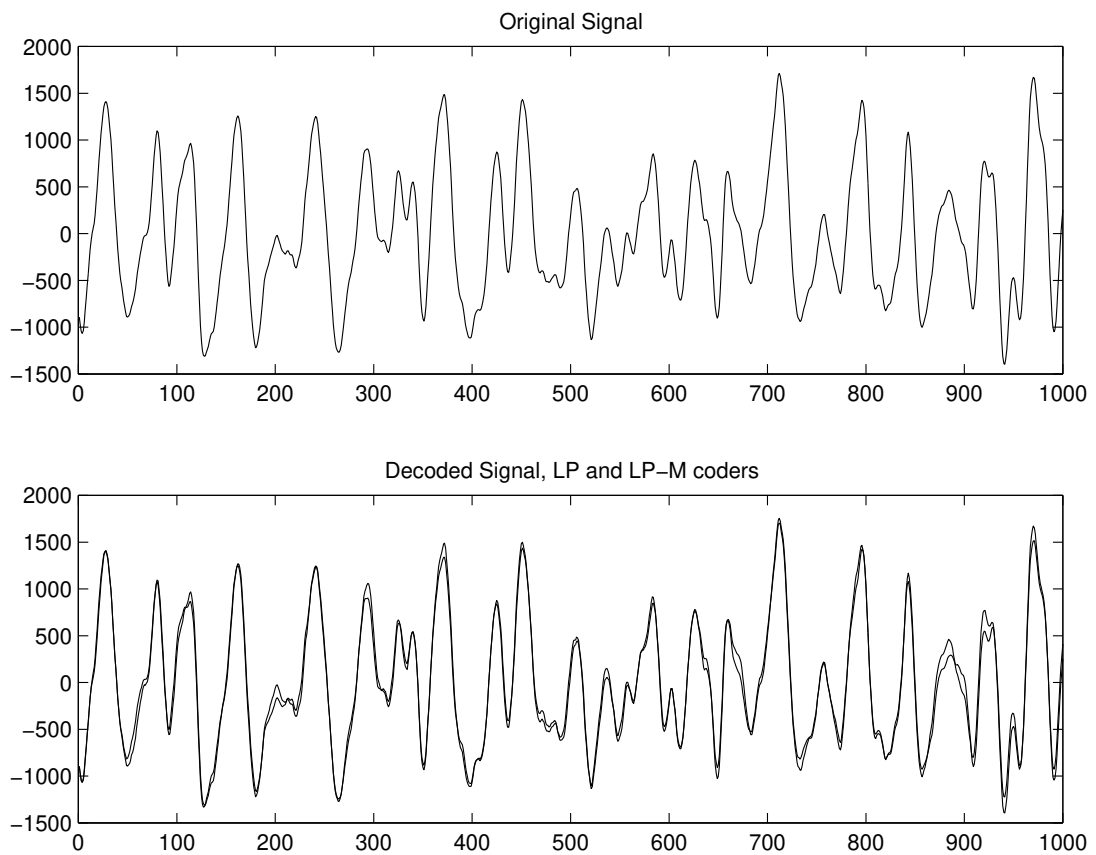


FIGURE 7.7: *AR Signal and CODEC approximation*

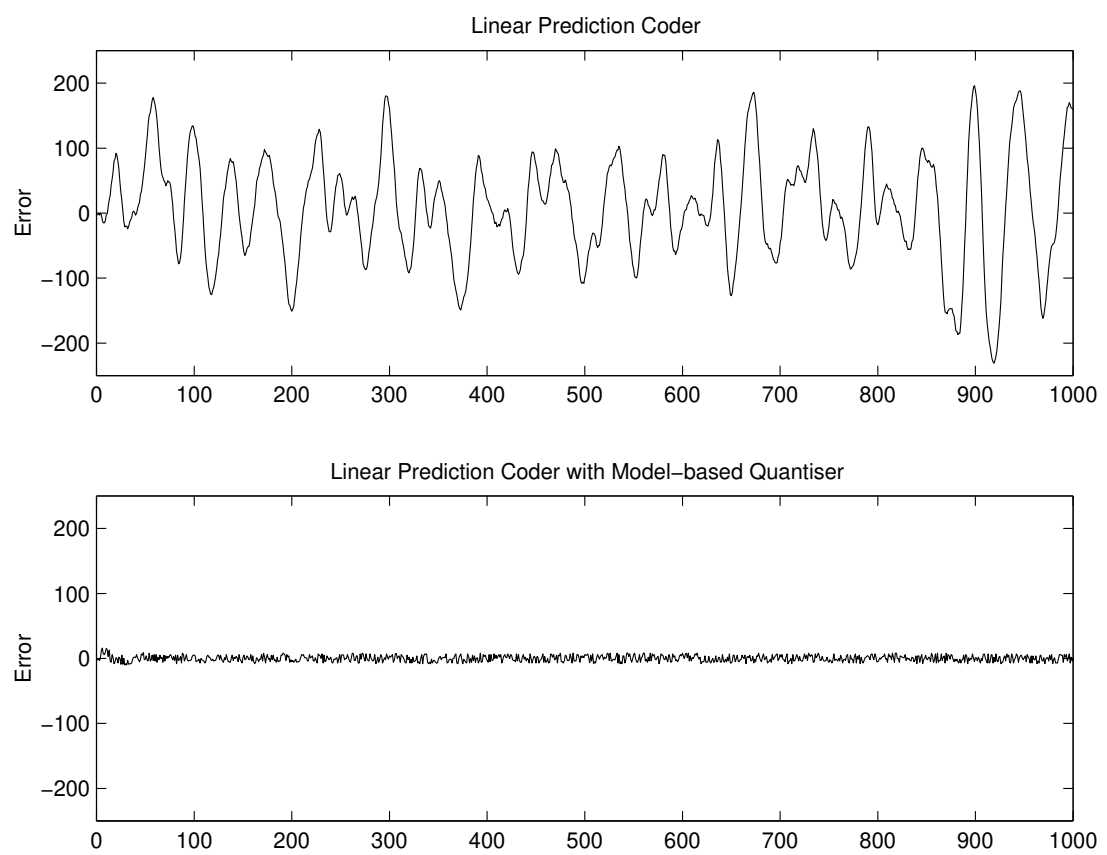


FIGURE 7.8: *CODEC Signal Reconstruction Error*

7.4.4 Coder Demonstration

Items [17]–[19] demonstrate the CODEC systems for high quality musical source material. Track [17] is the reference, taken directly from a modern digital recording.

Tracks [18] and [19] have been compressed 4:1 with the basic and enhanced coders respectively. The basic coder imparts a modulation on the signal spectral peaks (*i.e.* the strong tonal components of the signal) which gives them a “warbling” quality. This is the result of concentrating the error energy close to the tonal peaks.

The enhanced CODEC, heard on track [19], is of an audibly higher quality. The tonal modulation is not present, and the perceived noise floor is raised little over the original material.

7.5 Conclusions

In this chapter we have shown that quantisation of signal samples introduces errors which can be made to have any desired spectral shape by the addition of feedback around the quantiser. We have presented a new technique in which the filter in such a feedback loop is based on a signal model, and have termed systems of this type “model-based quantisers”.

Two applications were illustrated. The first showed how the system can be used to automatically move quantisation noise away from parts of the spectrum occupied by desired signal components. This is of use principally when a signal occupies a small proportion of the available bandwidth.

The second application showed how the model-based quantiser can be used to enhance the performance of a simple Linear Prediction CODEC. Its incorporation into the coder results in the reconstruction error being both whitened and reduced in amplitude. This is achieved with no modification to the decoder.

Conclusions

THIS DISSERTATION has analysed a number of models for multi-channel audio signals, and developed algorithms for estimation of their parameters. Their application to audio restoration and coding algorithms has been demonstrated. Additionally, methods for accurately resynchronising a number of audio signals have been developed.

8.1 Signal Models

Two models were considered for the case where multiple observations are available of a single underlying signal. This situation occurs, for example, when multiple copies of a disc or tape are available, or when multiple signals are extracted from a single carrier by means of a multi-channel head or pick-up.

In the Multiply-Observed AR (MO-AR) model, a single AR signal is degraded by a number of independent interference signals to give the multiple observations. This framework allows separation of the signal and noise statistics to a greater extent than the equivalent single-channel model. Under the assumption of Gaussianity and independence of the interference sources their autocorrelations and spectra may be estimated from the observed signals. These noise spectra may then be used as the basis for a signal subtraction or frequency domain Wiener filtering approach to broadband noise reduction.

The second model is the Ensemble-AR (E-AR) model. In this case we use an ensemble of excitation sources to drive identical AR filters. The outputs of these filters are the multiple observations. The parameters of this model may be efficiently estimated from the observed data. This model was used to develop a multi-channel system for detection and removal of impulsive noise. This system with just two channels is much superior to equivalent single-channel systems.

The Coupled-ARMA (C-ARMA) system was put forward as a possible model for stereo audio signals. Stereo signals comprise two sub-signals, generally referred to as “left” and “right”, and use these to generate the auditory illusion of spatially-separated sources. The C-ARMA model has a single recursive section, and two separate moving average sections. A white excitation signal drives the recursive section, and the output of this is applied to each of the moving average sections. The outputs of these are the left and right channels of the stereo signal.

Parameter estimation for the single-channel ARMA model is a difficult and computationally expensive problem. The structure of the C-ARMA system allows efficient estimation of all the parameters. Various interpolators for this for this system were derived, and all worked well on synthetic data. The single-channel ARMA interpolator was found to work satisfactorily given model parameters derived for the C-ARMA model. Stereo audio data appears qualitatively to fit the C-ARMA model, but the most powerful of the interpolation algorithms, the joint MAP interpolator, performs poorly on such audio data.

8.2 Synchronisation

It was found to be important that the various channels in a multi-channel audio system are accurately aligned. A method for measuring the time-offset between a pair of signals and compensating for it has been developed, and was demonstrated to be effective at aligning a pair of transcriptions from gramophone records.

8.3 Applications

The dissertation has been illustrated with engineering applications of audio signal models. These have principally been in audio restoration and noise reduction, where the multi-channel methods have significant advantages over equivalent single-channel methods. In particular, algorithms for multi-channel broadband

noise reduction, and multi-channel impulsive noise detection and removal have been demonstrated which outperform their single-channel counterparts.

In addition, a novel application of signal models was suggested, in which the behaviour of a quantiser can be made dependent on the modelled characteristics of the signal passing through it. Two applications were shown. Firstly the system was used to manipulate the noise floor of a quantised narrow band signal. Secondly, the system was used to enhance the performance of a simple linear prediction coder.

8.4 Further Research

8.4.1 Multi-channel Impulsive Noise Detection/Interpolation

The impulsive noise detection and removal system is essentially the multi-channel extension of work done by Vaseghi, Rayner and Godsill between 1987 and 1991. Since then many advances have been made in the single-channel system (see [39] for an overview and bibliography), and many of these could also benefit from a multi-channel approach.

For example, incorporation of the time-distribution statistics of impulsive degradation would allow it to take account of the “bursty” nature of such noise. This would be expected to yield greater robustness in the determination of exactly which samples are corrupt, particularly if incorporated in a Bayesian detection/interpolation scheme such as that presented in [39].

8.4.2 Statistical Signal Processing

There has been, with the increasing availability of computational power, an upsurge of interest in statistical sampling methods [34]. Such methods enable investigation of probability density functions which are intractable to analytic solution.

The parameter estimation problem for the MO-AR model was found to be a complicated p.d.f., but whose solution may well yield to a sampling approach. This would, in turn, allow the development of a model-based broadband noise reduction system based on that model.

Sampling techniques also facilitate the exploration of p.d.f.’s which are not so analytically convenient as the Gaussian. As a result of this it is sometimes possible to remove assumptions of Gaussianity, and closer approach the true distributions

of, for example, the impulsive noise found in many audio recordings.

8.4.3 General Audio Signal Research

8.4.3.1 Stereo Signals

It is felt, through this research, that a deeper investigation of the nature of stereo signals would be profitable. Stereo signals are so prevalent, yet, it seems, little understood. Few current audio coding algorithms make use of inter-channel redundancy at all, choosing instead to code the two signals independently. Those that do code stereo signals as a single entity use primitive sums and differences of the signals to obtain marginal increases in coding gain for the majority of signals. Such an investigation might also shed light on why the C-ARMA interpolator developed in chapter 6 performs poorly on audio signals.

Further to this, coders for multi-channel audio signals are becoming more important. Home cinema systems of the relatively near future will require high quality multi-channel audio to be delivered as part of video-on-demand and similar services. At the time of writing the coding gain of video coders is very much higher than that of audio coders.

8.4.3.2 One-bit Signals

An interesting area, but one which I feel is likely to remain something of a niche in terms of the audio industry at large, is the coding and processing of one-bit PCM audio signals. Most of the analogue to digital and digital to analogue converters used for digital audio are sigma-delta modulators which naturally produce a highly-oversampled single-bit representation of the audio. The high level of quantisation noise associated with a one-bit PCM signal is noise-shaped out of the audio passband by feedback around the quantiser.

The possibility of processing and coding this bit-stream directly is interesting, as it eliminates the need for the decimation and oversampling filters that are otherwise required for conversion to and from baseband PCM. A system that can mix a number of such signals, and apply some audio-band equalisation to them has been demonstrated by Sony Corporation. More complex processing of the signals represents a severe intellectual challenge.

Appendices

Demonstration CD

A

THE DEMONSTRATION CD accompanying this dissertation illustrates many of the techniques and algorithms presented in the text. The following is a listing and brief description of each track, with details of the source material where these are known. For full details of each track, and the phenomena which each demonstrates, the reader is referred to the main text.

The assistance of Mr E Kendall in researching the source material and making the transcriptions for items [1]–[4], [7], [9] and [10] is gratefully acknowledged.

Item [17] is copyright ©1997 of the Classical Recording Company, and is reproduced with permission.

The CD is not included with this copy of the dissertation; please refer to the Signal Processing Group website (<http://www-sigproc.eng.cam.ac.uk>) for further details.

TRACK	SEC.	DETAILS	DESCRIPTION
<div style="display: flex; flex-direction: column; gap: 5px;"> <div style="display: flex; gap: 5px;"> 1 2 </div> <div style="display: flex; gap: 5px;"> 3 4 </div> 5 6 </div>	3.4.1	QHCF, <i>Minor Swing</i> , OLA 1990-1 (1937)	<p>Two-channel transcriptions of four copies of a single pressing.</p> <p>Eight-channel restoration of 1-4 using spectral subtraction.</p> <p>Eight-channel restoration using marginalised p.d.f.</p>
<div style="display: flex; flex-direction: column; gap: 5px;"> 7 8 </div>	4.5.5	It don't mean a thing, Stéphane Grappelli, Polydor 2083 HPP (1935)	<p>Monophonic disc showing independent impulsive noise in the two channels of this transcription.</p> <p>7 restored using two-channel impulsive noise detector and interpolator.</p>
<div style="display: flex; flex-direction: column; gap: 5px;"> <div style="display: flex; gap: 5px;"> 9 10 </div> 11 12 </div>	6.2.1	Intermezzo from Symphony No. 10, Mahler, SPA31 (c.1955)	<p>Two-channel transcriptions of two copies of a monophonic LP.</p> <p>Sum of 9 and 10 exhibits comb-filtering due time offset.</p> <p>Sum of 9 and 10 after resynchronisation.</p>
<div style="display: flex; flex-direction: column; gap: 5px;"> 13 14 15 16 </div>	7.3.3	<i>Synthetic</i>	<p>Narrowband two-tone test signal.</p> <p>13 quantised conventionally.</p> <p>13 quantised with model-based quantiser, moving quantisation noise away from spectral peaks.</p> <p>13 quantised with model-based quantiser, moving quantisation noise towards spectral peaks.</p>
<div style="display: flex; flex-direction: column; gap: 5px;"> 17 18 19 </div>	7.4.4	Hodie Christus Natus Est, Rihards Dubra, CRC701-2 (1997)	<p>High quality stereophonic test material</p> <p>17 compressed approx. 4:1 using conventional LPC.</p> <p>17 compressed approx. 4:1 using enhanced LPC.</p>

B

Correlation Calculations

IT IS WELL KNOWN that the auto-correlation of a random signal and its power spectral density are related by the Fourier transform. Similarly the cross-spectrum of a pair of signals is the Fourier transform of their cross-correlation. Thus, correlation functions for discrete-time signals may be efficiently estimated *via* the Fast Fourier Transform (FFT) algorithm. The method is outlined here, as details of the implementation are scarce in the literature.

The discrete cross-correlation of a pair of random signals $x[n]$ and $y[n]$ is given by

$$\mathcal{R}_{xy}[\tau] = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=-\frac{N}{2}}^{\frac{N}{2}-1} x[n] y[n - \tau] \quad (\text{B.1})$$

where τ is the *lag*.

We describe here one specific, straightforward algorithm which calculates the function

$$\mathcal{R}_{xy}[\tau] = \frac{1}{N} \sum_{n=-\frac{N}{2}}^{\frac{N}{2}-1} x[n] y[n - \tau], \quad -N < \tau < N \quad (\text{B.2})$$

as an approximation to the true cross-correlation, where $x[n]$ and $y[n]$ are both zero outside the range $-\frac{N}{2} \leq n < \frac{N}{2}$.

The method may be trivially adapted to calculation of the auto-correlation function by the substitution $y[n] = x[n]$. It should be noted that $R_{xy}[\tau]$ is zero outside the range $-N < \tau < N$.

B.1 Efficient Estimation of the Cross-correlation Function

We start by generating zero-padded signals $x'[n]$ and $y'[n]$, each of length $2N$, by adding $\frac{N}{2}$ zeros at each end of each of x and y . These padded sequences therefore begin at index $n = -N$ and end at index $n = N - 1$.

Let $X[k]$ and $Y[k]$ be the Discrete Fourier Transforms (evaluated *via* the FFT) of these padded sequences, given by

$$X[k] = \sum_{m=-N}^{N-1} x'[m] \exp\left(\frac{-j2\pi mk}{2N}\right) \quad (\text{B.3})$$

$$Y[k] = \sum_{n=-N}^{N-1} y'[n] \exp\left(\frac{-j2\pi nk}{2N}\right) \quad (\text{B.4})$$

for $-N \leq k \leq N - 1$.

The product $S_{xy}[k] = X^*[k]Y[k]$ is known as the cross-spectrum. Taking the inverse FFT of this product we obtain

$$\frac{1}{2N} \sum_{k=-N}^{N-1} \left[\sum_{m=-N}^{N-1} x'[m] \exp\left(\frac{j\pi mk}{N}\right) \sum_{n=-N}^{N-1} y'[n] \exp\left(\frac{-j\pi nk}{N}\right) \right] \exp\left(\frac{j\pi k\tau}{N}\right) \quad (\text{B.5})$$

which can be rearranged

$$\frac{1}{2N} \sum_{m=-N}^{N-1} \sum_{n=-N}^{N-1} \left[x'[m] y'[n] \sum_{k=-N}^{N-1} \exp\left(\frac{-j\pi k}{N}(\tau + m - n)\right) \right] \quad (\text{B.6})$$

by swapping the order of the summations.

By orthogonality, the complex exponential in equation B.6 sums to zero except when $\tau + m - n = 2pN$ for integer p . Thus multiplying expression B.6 by two we obtain

$$\frac{1}{N} \sum_{m=-N}^{N-1} \sum_{n=-N}^{N-1} x'[m] y'[n] \delta[(\tau + m - n) \circ 2N] \quad (\text{B.7})$$

where \circ denotes the modulo operator.

The discrete delta function $\delta(\cdot)$ selects only the summation terms for which $\tau = (n - m) \circ 2N$. Thus expression B.7 for $-N \leq \tau < N$ evaluates identically to expression B.2.

B.1.0.3 Summary

The cross-correlation function calculation may be summarised as follows:

- zero-pad the sequences $x[n]$, $y[n]$ to length $2N$,
- calculate the FFT of both sequences, $X[k]$, $Y[k]$,
- calculate the product $S_{xy}(k) = X^*[k]Y[k]$,
- take the IFFT of S_{xy} and divide by 2 to give $R_{xy}[\tau]$.

Beware that there is little consensus as to the indexing associated with FFT algorithms; for an N -point FFT, some sources assume $0 \leq n < N$, whereas others take $-\frac{N}{2} \leq n < \frac{N}{2}$ as we have done here.

Partial Cross-Correlation Function

Frequently we are interested only in lags close to $\tau = 0$. If we require lags up to only $\tau = \pm(P - 1)$ then there are small additional savings to be made by splitting the data sequences into sub-sequences of P samples each. The correlation is then calculated in terms of the correlations of the sub-sequences, each of which is calculated by the method above.

B.1.1 Computational Considerations

The computational requirement for this algorithm splits down as follows:

- two FFTs, each of length $2N$,
- $2N$ complex multiplications,
- one inverse FFT of length $2N$.

The number of multiply-accumulate (MAC) instructions required for a $2N$ point (inverse) FFT is approximately $4N \log_2(2N)$, so the total requirement for this algorithm is approximately $N(8 + 12 \log_2(2N))$ instructions.

This compares with approximately N^2 instructions for direct evaluation of equation B.2. The exact data set size at which the FFT-based method becomes more efficient than direct calculation depends on the detail of the DSP architecture and instruction set.

Solution of the inequality based purely on the arithmetic calculations suggests that it is beneficial to use the FFT-based algorithm for $N \geq 128$.

C

Integrals and the Gaussian PDF

THE GAUSSIAN P.D.F. for a real vector \underline{u} is given by

$$p_{\underline{u}}(\underline{u}) = \frac{1}{(2\pi)^{N/2} |\mathbf{R}_{\underline{u}}|^{1/2}} \exp\left(-\frac{(\underline{u} - \underline{m}_{\underline{u}})^T \mathbf{R}_{\underline{u}}^{-1} (\underline{u} - \underline{m}_{\underline{u}})}{2}\right) \quad (\text{C.1})$$

where \underline{u} is of length N , with mean $\underline{m}_{\underline{u}}$ and covariance matrix $\mathbf{R}_{\underline{u}}$.

Integrals of the form

$$I = \int_{\mathcal{R}^N} \exp\left(-\frac{1}{2} (\underline{x}^T \mathbf{A}^T \mathbf{A} \underline{x} + \underline{s}^T \underline{x} + d)\right) d\underline{x} \quad (\text{C.2})$$

appear at a number of places in the dissertation. The infinitesimal volume element $d\underline{x}$ is interpreted as

$$d\underline{x} = \prod_{n=1}^N dx_n \quad (\text{C.3})$$

and the integration is performed over the infinite N -dimensional real space \mathcal{R}^N . In other words the symbol $\int_{\mathcal{R}^N}$ is interpreted as

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}$$

The result

$$I = \left(\frac{(2\pi)^N}{|\mathbf{A}^T \mathbf{A}|}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{2} \left(d - \frac{\underline{s}^T (\mathbf{A}^T \mathbf{A})^{-1} \underline{s}}{4}\right)\right) \quad (\text{C.4})$$

may be derived by completing the square

$$\underline{\mathbf{x}}^T \mathbf{A}^T \mathbf{A} \underline{\mathbf{x}} + \underline{\mathbf{s}}^T \underline{\mathbf{x}} + \mathbf{d} = (\underline{\mathbf{x}} - \underline{\mathbf{m}}_x)^T \mathbf{A}^T \mathbf{A} (\underline{\mathbf{x}} - \underline{\mathbf{m}}_x) + \mathbf{k} \quad (\text{C.5})$$

where

$$\mathbf{k} = \left(\mathbf{d} - \frac{\underline{\mathbf{s}}^T (\mathbf{A}^T \mathbf{A})^{-1} \underline{\mathbf{s}}}{4} \right) \quad (\text{C.6})$$

$$\underline{\mathbf{m}}_x = -\frac{(\mathbf{A}^T \mathbf{A})^{-1} \underline{\mathbf{s}}}{2}. \quad (\text{C.7})$$

We may now rewrite the integral as

$$\mathbf{I} = \int_{\mathcal{R}^N} \exp \left(-\frac{1}{2} ((\underline{\mathbf{x}} - \underline{\mathbf{m}}_x)^T \mathbf{A}^T \mathbf{A} (\underline{\mathbf{x}} - \underline{\mathbf{m}}_x) + \mathbf{K}) \right) \mathrm{d}\underline{\mathbf{x}} \quad (\text{C.8})$$

$$= \exp \left(-\frac{1}{2} \mathbf{K} \right) \int_{\mathcal{R}^N} \exp \left(-\frac{(\underline{\mathbf{x}} - \underline{\mathbf{m}}_x)^T \mathbf{A}^T \mathbf{A} (\underline{\mathbf{x}} - \underline{\mathbf{m}}_x)}{2} \right) \mathrm{d}\underline{\mathbf{x}}. \quad (\text{C.9})$$

By comparison with the Gaussian p.d.f. (equation C.1) which has unit volume we may determine that

$$\mathbf{I} = \left(\frac{(2\pi)^N}{|\mathbf{A}^T \mathbf{A}|} \right)^{\frac{1}{2}} \exp \left(-\frac{1}{2} \left(\mathbf{d} - \frac{\underline{\mathbf{s}}^T (\mathbf{A}^T \mathbf{A})^{-1} \underline{\mathbf{s}}}{4} \right) \right) \quad (\text{C.10})$$

D

MO-AR Model Error Variances

IN THIS appendix we derive the results quoted in chapter 3 concerning the expected errors associated with various signal estimates.

D.1 Weighted Estimate Error Power

The weighted signal estimate for the multiple additive noise source model was given as

$$\hat{u}_x[n] = \left(\sum_{q=1}^Q \frac{1}{\sigma_{n_q}^2} \right)^{-1} \sum_{q=1}^Q \frac{x_q[n]}{\sigma_{n_q}^2}. \quad (\text{D.1})$$

Substituting $x_q[n] = u[n] + n_q[n]$ and

$$K = \sum_{q=1}^Q \frac{1}{\sigma_{n_q}^2}$$

and dropping the time index $[n]$ for clarity gives

$$\hat{u}_x = \frac{1}{K} \sum_{q=1}^Q \frac{u + n_q}{\sigma_{n_q}^2}. \quad (\text{D.2})$$

The error power is given by

$$\mathbb{E} [(u - \hat{u}_x)^2] = \mathbb{E} \left[\left(u - \frac{1}{K} \sum_{q=1}^Q \frac{u + n_q}{\sigma_{n_q}^2} \right)^2 \right] \quad (\text{D.3})$$

$$= \frac{1}{K^2} \mathbb{E} \left[\left(Ku - \sum_{q=1}^Q \frac{u + n_q}{\sigma_{n_q}^2} \right)^2 \right] \quad (\text{D.4})$$

$$= \frac{1}{K^2} \mathbb{E} \left[\left(Ku - u \sum_{q=1}^Q \frac{1}{\sigma_{n_q}^2} - \sum_{q=1}^Q \frac{n_q}{\sigma_{n_q}^2} \right)^2 \right] \quad (\text{D.5})$$

$$= \frac{1}{K^2} \mathbb{E} \left[\left(Ku - Ku - \sum_{q=1}^Q \frac{n_q}{\sigma_{n_q}^2} \right)^2 \right] \quad (\text{D.6})$$

$$= \frac{1}{K^2} \mathbb{E} \left[\sum_{q_1=1}^Q \frac{n_{q_1}}{\sigma_{n_{q_1}}^2} \sum_{q_2=1}^Q \frac{n_{q_2}}{\sigma_{n_{q_2}}^2} \right] \quad (\text{D.7})$$

The expectation of every product term is zero, except for those where $q_1 = q_2$; we may therefore write

$$\mathbb{E} [(u - \hat{u}_x)^2] = \frac{1}{K^2} \sum_{q=1}^Q \mathbb{E} \left[\frac{n_q^2}{\sigma_{n_q}^4} \right] \quad (\text{D.8})$$

$$= \frac{1}{K^2} \sum_{q=1}^Q \frac{\sigma_{n_q}^2}{\sigma_{n_q}^4} \quad (\text{D.9})$$

$$= \frac{1}{K^2} K \quad (\text{D.10})$$

$$= \left(\sum_{q=1}^Q \frac{1}{\sigma_{n_q}^2} \right)^{-1} \quad (\text{D.11})$$

as quoted in equation 3.23.

D.2 Unweighted Estimate Error Power

The unweighted signal estimate $\bar{x}[\mathbf{n}]$ is given by

$$\bar{x}[\mathbf{n}] = \frac{1}{Q} \sum_{q=1}^Q x_q[\mathbf{n}] \quad (\text{D.12})$$

Dropping the time index, and substituting $x_q = u + n_q$ gives the estimation error power

$$\mathbb{E} \left[(u - \bar{x})^2 \right] = \mathbb{E} \left[\left(u - \frac{1}{Q} \sum_{q=1}^Q (u + n_q) \right)^2 \right] \quad (\text{D.13})$$

$$= \mathbb{E} \left[\left(u - u - \frac{1}{Q} \sum_{q=1}^Q n_q \right)^2 \right] \quad (\text{D.14})$$

$$= \frac{1}{Q^2} \mathbb{E} \left[\left(\sum_{q_1=1}^Q n_{q_1} \right) \left(\sum_{q_2=1}^Q n_{q_2} \right) \right] \quad (\text{D.15})$$

$$= \frac{1}{Q^2} \sum_{q=1}^Q \mathbb{E} [n_q^2] \quad (\text{D.16})$$

$$= \frac{1}{Q^2} \sum_{q=1}^Q \sigma_{n_q}^2 \quad (\text{D.17})$$

as quoted in equation 3.26.

D.3 Comparison of weighted and unweighted signal estimates

We wish to prove that

$$10 \log_{10} \left(\sigma_u^2 \sum_{q=1}^Q \frac{1}{\sigma_{n_q}^2} \right) \geq 10 \log_{10} \left(\sigma_u^2 \left(\frac{1}{Q^2} \sum_{q=1}^Q \sigma_{n_q}^2 \right)^{-1} \right) \quad (\text{D.18})$$

as was asserted in equation 3.28. In order to do this it will be adequate to show that the noise powers observe the relationship

$$\mathbb{E} \left[(u - \bar{x})^2 \right] \geq \mathbb{E} \left[(u - \hat{u}_{\bar{x}})^2 \right] \quad (\text{D.19})$$

since the signal power in each case is identical.

Rewriting the expectations as a ratio gives

$$\frac{\mathbb{E} \left[(u - \bar{x})^2 \right]}{\mathbb{E} \left[(u - \hat{u}_{\bar{x}})^2 \right]} = \frac{1}{Q^2} \left(\sum_{q=1}^Q \sigma_{n_q}^2 \right) \left(\sum_{q=1}^Q \frac{1}{\sigma_{n_q}^2} \right) \quad (\text{D.20})$$

$$= \frac{1}{Q^2} \sum_{q_1=1}^Q \sum_{q_2=1}^Q \frac{\sigma_{n_{q_1}}^2}{\sigma_{n_{q_2}}^2} \quad (\text{D.21})$$

$$= \frac{1}{Q^2} \left[\sum_{q=1}^Q \frac{\sigma_{n_q}^2}{\sigma_{n_q}^2} + \sum_{q_1=1}^Q \sum_{q_2=q_1+1}^Q \left(\frac{\sigma_{n_{q_1}}^2}{\sigma_{n_{q_2}}^2} + \frac{\sigma_{n_{q_2}}^2}{\sigma_{n_{q_1}}^2} \right) \right] \quad (\text{D.22})$$

Since $\frac{a}{b} + \frac{b}{a} \geq 2$ we may write

$$\frac{E[(u - \bar{x})^2]}{E[(u - \hat{u}_x)^2]} \geq \frac{1}{Q^2} \left[Q + \sum_{q_1=1}^Q \sum_{q_2=q_1+1}^Q 2 \right] \quad (\text{D.23})$$

$$\geq \frac{Q + 2T_{(Q-1)}}{Q^2} \quad (\text{D.24})$$

where T_N is the N^{th} triangular number

$$T_N = \frac{N(N+1)}{2} \quad (\text{D.25})$$

Substituting D.25 into D.24 gives

$$\frac{E[(u - \bar{x})^2]}{E[(u - \hat{u}_x)^2]} \geq \frac{Q + (Q-1)Q}{Q^2} \quad (\text{D.26})$$

$$\geq \frac{Q^2}{Q^2} \quad (\text{D.27})$$

$$\geq 1 \quad (\text{D.28})$$

Least Squares and Associated Algorithms

E

THE TOTAL LEAST SQUARES method [42] provides an alternative solution for parameter estimation problems, in which the assumptions made about the errors in the observed data differ from those made in the ordinary least squares method.

E.1 Ordinary Least Squares Method

The least squares method provides solutions to systems of the form

$$X\underline{\mathbf{b}} = \underline{\mathbf{d}} - \underline{\mathbf{e}} \quad (\text{E.1})$$

where X and $\underline{\mathbf{d}}$ are known, and we wish to find $\underline{\mathbf{b}}$ such that the error $\underline{\mathbf{e}}^T \underline{\mathbf{e}}$ is minimised. The solution $\underline{\mathbf{b}}_{\text{LS}}$ is well-known and given by

$$(X^T X) \underline{\mathbf{b}}_{\text{LS}} = X^T \underline{\mathbf{d}} \quad (\text{E.2})$$

assuming that the system is not rank-deficient.

E.2 Approximate Least Squares

In some cases the matrix $X^T X$ is approximately Toeplitz; that is, each of the diagonals of matrix $X^T X$ contains elements which are approximately equal. This situation occurs, for example, in the estimation of AR model coefficients.

If an appropriate Toeplitz approximation can be found for $X^T X$ then an approximate LS solution may be found very efficiently using Levinson recursion [61, 97].

E.3 Total Least Squares Method

The explanation of the TLS algorithm presented here is based on that given by Therrien in [97]. The principal simplification over the more general method presented by Golub and Van Loan [42], is that the latter is not limited to vector \underline{b} and \underline{d} .

The form of equation E.1 implicitly associates the errors with vector \underline{d} . However, it is often the case that both X and \underline{d} contain measured experimental data, and as a result that both are subject to observation noise. It therefore would seem desirable to re-formulate the problem as

$$(X - E) \underline{b} = \underline{d} - \underline{e} \quad (\text{E.3})$$

and to find a solution which, in some sense, minimises both \underline{e} and E . The method of ‘‘Total Least Squares’’ provides one such solution.

Let us suppose, for the following discussion, that matrix X has K rows and P columns, and that $K \geq P + 1$.

Equation E.3 may be rearranged as follows

$$(X - E) \underline{b} - \underline{d} + \underline{e} = \underline{0} \quad (\text{E.4})$$

$$\left[(X - E) \mid (\underline{d} - \underline{e}) \right] \begin{bmatrix} \underline{b} \\ -1 \end{bmatrix} = \underline{0} \quad (\text{E.5})$$

$$\left([X \mid \underline{d}] - [E \mid \underline{e}] \right) \begin{bmatrix} \underline{b} \\ -1 \end{bmatrix} = \underline{0} \quad (\text{E.6})$$

$$(M - W) \begin{bmatrix} \underline{b} \\ -1 \end{bmatrix} = \underline{0} \quad (\text{E.7})$$

where

$$M = \left[X \mid \underline{d} \right] \quad (\text{E.8})$$

$$W = \left[E \mid \underline{e} \right] \quad (\text{E.9})$$

Note that equation E.7 implies that the matrix $(M - W)$ is, by definition, rank-deficient.

The root sum of the squares of the elements of a matrix M

$$\|M\|_F = \left(\sum_{i=1}^K \sum_{j=1}^{P+1} M_{i,j}^2 \right)^{\frac{1}{2}} \quad (\text{E.10})$$

is known as the Frobenius norm. It may be shown that the squared Frobenius norm is equal to the sum of the squared singular values.

$$M = \sum_{i=1}^{P+1} \sigma_i \underline{u}_i \underline{v}_i^T \quad (\text{E.11})$$

$$\|M\|_F^2 = \sum_{i=1}^{P+1} \sigma_i^2 \quad (\text{E.12})$$

The Total Least Squares method calculates a matrix W_0 such that equation E.7 is satisfied for some \underline{b} , and such that the squared Frobenius norm of W_0 is minimised.

The SVD of M is given by equation E.11 and it can be shown that the matrix W_0 of smallest squared Frobenius norm that makes $(M - W_0)$ rank-deficient is given by

$$W_0 = \sigma_{P+1} \underline{u}_{P+1} \underline{v}_{P+1}^T \quad (\text{E.13})$$

where σ_{P+1} is the smallest singular value of M .

In order to find the solution $\underline{b}_{\text{TLS}}$ we substitute W_0 from E.13 into equation E.7 to give

$$(M - \sigma_{P+1} \underline{u}_{P+1} \underline{v}_{P+1}^T) \begin{bmatrix} \underline{b}_{\text{TLS}} \\ -1 \end{bmatrix} = \underline{0} \quad (\text{E.14})$$

$$\left(\sum_{i=1}^P \sigma_i \underline{u}_i \underline{v}_i^T \right) \begin{bmatrix} \underline{b}_{\text{TLS}} \\ -1 \end{bmatrix} = \underline{0} \quad (\text{E.15})$$

For this condition to be satisfied it is clear that $[\underline{b}_{\text{TLS}}^T \mid -1]^T$ must be proportional to \underline{v}_{P+1} , since it is required by equation E.15 to be orthogonal to all of \underline{v}_i , $i = 1 \dots P$. Hence

$$\begin{bmatrix} \underline{b}_{\text{TLS}} \\ -1 \end{bmatrix} = c \underline{v}_{P+1} \quad (\text{E.16})$$

and by partitioning \underline{v}_{P+1}

$$\begin{bmatrix} \underline{b}_{\text{TLS}} \\ -1 \end{bmatrix} = c \begin{bmatrix} \underline{v} \\ v \end{bmatrix} \quad (\text{E.17})$$

Solving the bottom elemental equation in E.17 gives $\mathbf{c} = -1/v$, and hence

$$\underline{\mathbf{b}}_{\text{TLS}} = -\frac{1}{v} \underline{\mathbf{v}} \quad (\text{E.18})$$

There are two degeneracies associated with the TLS method.

- The TLS solution does not exist if \mathbf{M} is rank-deficient.
- If the smallest singular value has multiplicity $Q > 1$ then there are Q possible solutions.

In the former case it is possible that $\mathbf{X}^T \mathbf{X}$ is of full rank, in which case the ordinary least squares solution given by E.2 may be used. In the latter case it has been suggested that the solution with smallest Euclidean norm be selected [42].

E.4 Computational Considerations

The TLS method is considerably more computationally expensive than the ordinary LS method.

The computation for the LS algorithm is dominated by the matrix inverse of the $\mathbf{P} \times \mathbf{P}$ matrix $\mathbf{X}^T \mathbf{X}$. This operation is order $\mathcal{O}(\mathbf{P}^3)$. If the problem is such that the Toeplitz approximation may be made then an approximate LS solution may be found in $\mathcal{O}(\mathbf{P}^2)$ operations using Levinson recursion [61, 77]. For very large Toeplitz systems there exist algorithms of order $\mathbf{P}(\log(\mathbf{P}))^2$ but these are highly memory intensive [22].

By contrast, the TLS algorithm requires the calculation of the SVD of a $(\mathbf{P}+1) \times \mathbf{N}$ matrix. For $\mathbf{N} \geq (\mathbf{P}+1)$ this operation is order $\mathcal{O}((\mathbf{P}+1)\mathbf{N}^2 + \mathbf{N}^3)$ [42]. Although this cubic *order* is no worse than the LS algorithm we typically have $\mathbf{N} \gg \mathbf{P}$ in this application, and the number of calculations required for the TLS method is vastly greater.

Figure G.8 shows the logarithm (base 10) of the number of calculations (counted using MATLAB) for each of the algorithms (LS, Toeplitz approximation to LS, TLS) for a range of typical problem sizes. It is clear from the figure that the TLS method requires orders of magnitude more computation than LS or the Toeplitz approximation to LS for this range of problem sizes.

Code for both the Levinson recursion and for the SVD are given in [77].

F

Resampling of a Sampled Signal

In chapter 6 we examined methods for determining the time offset between a pair of signals which convey the same audio information, but which have distorted or degraded by independent mechanisms. Various techniques were examined for the determination of the time-axis warping function which maps the time axis of one signal to the time axis of the second.

Once the relative time axis warping function is known, one of the signals may be aligned with the other by use of a variable time shifter, set by the offset function derived during the detection phase. The shifter will need to be of sub-sample accuracy, and as such may be implemented as a polyphase filter similar to those used for sample rate conversion.

We have chosen a polyphase filterbank interpolator since it is simple to implement, and has adequate performance in this application. There are other interpolation algorithms in use for interpolation of sampled audio signals; a useful overview is given by Zolzer [108], where polynomial, Lagrange and spline interpolators are compared.

F.1 Filter Design

Suppose we wish to shift the signal $x[n]$ by the (non-integer) number of samples τ . The discrete impulse response $h_\tau[i]$ for a pure timeshift τ is given by

$$h_\tau[i] = \text{sinc}(\pi(i - \tau)), \quad -\infty < i < +\infty \quad (\text{F.1})$$

which represents a perfect bandpass filter whose amplitude response is unity across the whole Nyquist passband $-\frac{\pi}{2} < \omega < \frac{\pi}{2}$ and zero elsewhere. The shifted signal $x_\tau[n]$ is given by

$$x_\tau[n] = \sum_{i=-\infty}^{+\infty} h_\tau[i] x[n - i]. \quad (\text{F.2})$$

Note that an integer shift is a straightforward special case of this filter. For integer τ we obtain

$$h_\tau[i] = \begin{cases} 1, & i = \tau \\ 0, & \text{otherwise.} \end{cases} \quad (\text{F.3})$$

and this accords with our intuition that a shift of an integer number of samples is achieved simply by re-indexing the signal samples.

We clearly cannot implement equation F.2 directly as it requires an infinite summation for non-integer τ . The solution is to design a filter with a finite number of terms whose response approximates a pure time shift.

Simply windowing the impulse response F.1 is not usually a satisfactory approach since this introduces a finite transition band and significant aliasing results. It is usually preferable to design a low-pass filter with a cut-off slightly lower than the Nyquist bandwidth and then to window this suitably to obtain a finite set of filter coefficients.

In order to apply an integer shift in a system which uses windowed low-pass filters, we are required to filter the signal with the filter corresponding to $\tau = 0$. Simple re-indexing of the signal does not have the low-pass filtering effect and signal inconsistencies could result if it is not incorporated.

The choice of filter length, bandwidth and window will be dependent upon the application. The examples on the CD are of high quality musical recordings, and for these the filter cutoff is $0.9\frac{\pi}{2}$. A Hanning window was used to generate a filter of length 61 samples.

Where computation is at a premium, and where the signals are somewhat over-sampled, Rossum [86] gives an innovative technique for the design of low-order, high-performance interpolation filters. A 7-tap FIR filter with >120 dB attenuation in regions close to all multiples of the sample rate is demonstrated.

F.2 Efficient Implementation

The nature of this problem implies that we need to calculate a set of windowed filter coefficients for each value of τ that is encountered. We can, however, limit our consideration to shifts in the range $-\frac{1}{2} \leq \tau < \frac{1}{2}$ since larger shifts may be implemented as the superposition of an integer shift and a fractional shift in this range.

It is computationally expensive to calculate online the filter coefficients for each value of τ that may be encountered during the operation of a practical system. A convenient solution is to choose a desired time-shift resolution and pre-calculate the filters required to meet that resolution specification. For example, if we wish to be able to shift the signal with a resolution of 0.1 samples then we may pre-calculate windowed filters for $\tau = \{-0.5, -0.4 \dots, 0.3, 0.4\}$ samples and store the coefficients in a table. It is then a simple matter to choose the filter from the table which gives the closest time shift to the desired value at a given instant.

For typical DSP chip architectures and instruction sets it is at filter lengths greater than 64 or 128 samples that FFT methods are more efficient than direct calculation. If an FFT method is chosen, the transforms of the filters can, of course, be stored in the table rather than the coefficients themselves.

Variable Time Shift

The correction of varying time offsets clearly required that we apply a non-constant time shift to one of the signals. The offsets in the present applications vary sufficiently slowly that signal discontinuities rarely result from varying the time shift.

In other applications the shift may be faster than this, and in this case measures must be taken to ensure that the output signal is free of audible artifacts. For example, if FFT methods are employed to implement the filters, some block overlap will help to smooth the transitions between filters. It may also be desirable to limit the time-shift slew rate, such that effects such as an audible pitch shift

cannot occur.

Relationship to Sample Rate Conversion

Many sample rate conversion algorithms are based around similar resampling methods that use a windowed low-pass filter [20, 80]. For a fixed, rational sample rate ratio, the set of filters required is finite. These filters are applied to the incoming data samples in a deterministic sequence which mirrors the rotating phase relationship between the input and output samples at their respective rates. In this application the set of stored filters is often referred to as a *polyphase filterbank*.

The present application differs from this “synchronous” sample rate converter in that the effective instantaneous sample rate ratio is neither fixed nor necessarily rational. This situation is close to the “asynchronous” sample rate conversion problem [2, 3, 108], which ideally requires the calculation of a set of dedicated filter coefficients for each output sample. The calculated filter then corresponds to the precise instantaneous phase relationship between the input and output samples.

Online Correction

If the signal sources are free-running then it is difficult to apply the time-shift correction in the general case because there may be an average speed discrepancy between them. If the reference is the faster of the two signals we will potentially require infinite memory; if the reference is the slower, the channel to be shifted will have passed by us before we know what timeshift to apply to it.

The problem may in both cases be resolved by use of a signal delay long enough to contain the entire signal of interest. The most practical approach to implementing this delay is to transcribe the signals independently and store them on a computer in separate files. The starts may then be readily aligned to sufficient accuracy using an audio editor. It is then straightforward to read through each file at the required rate; we are guaranteed by the random-access nature of computer files always to have access to the required samples.

It may be possible, alternatively, to use the detected time-shift to control, *via* a suitable feedback system, the playback speed of one of the sources. While such a system would be academically pleasing it is not felt to be sufficiently practical for the present application to warrant further investigation here.

Colour Figures

G

THIS PAGE INTENTIONALLY LEFT BLANK

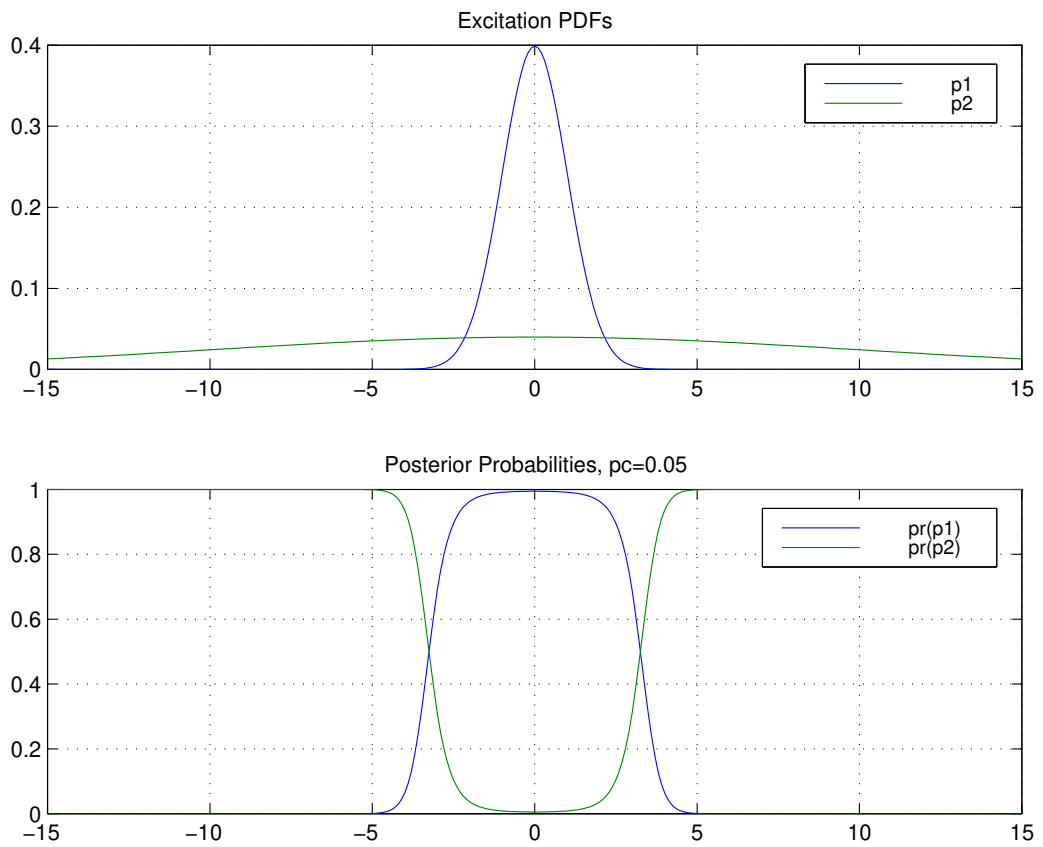


FIGURE G.1: *Excitation sample p.d.f's and posterior probabilities (section 4.3.3)*

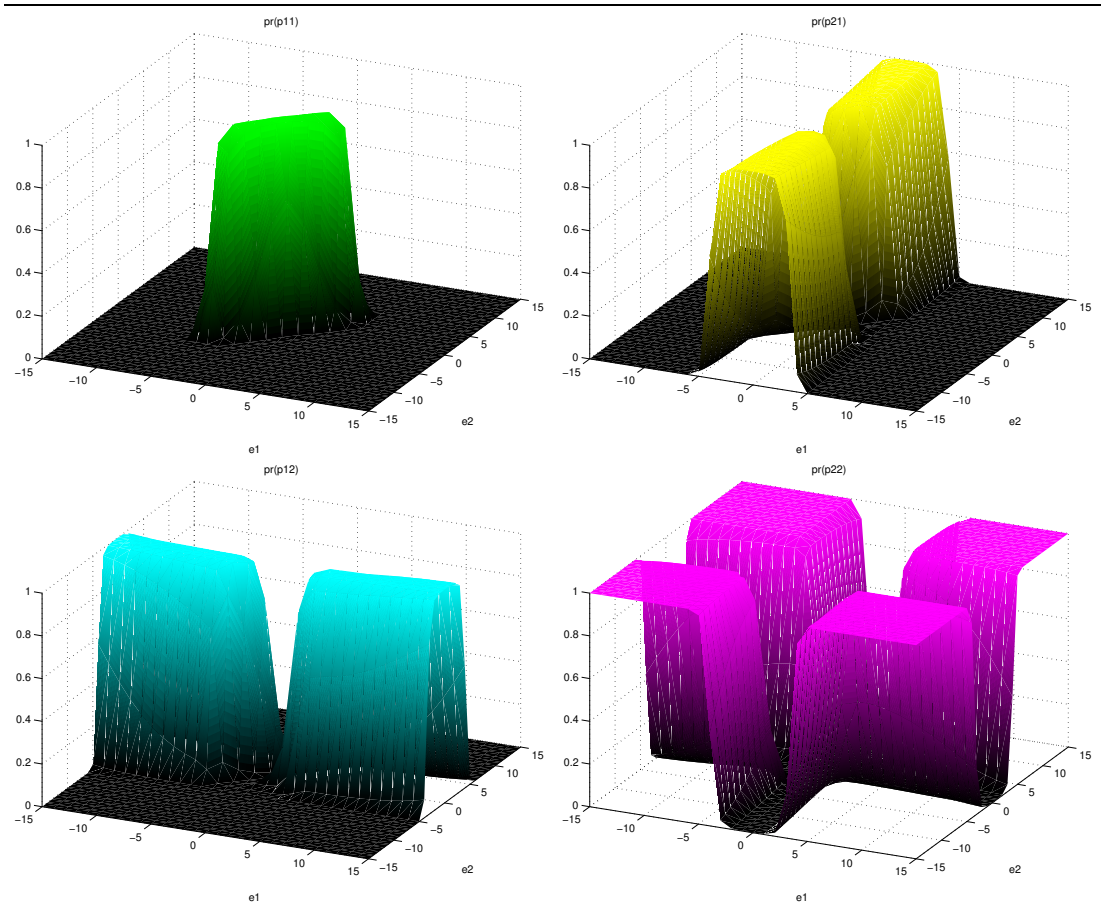


FIGURE G.2: Two-Channel E-AR Detector Posterior Probabilities (section 4.4.2)

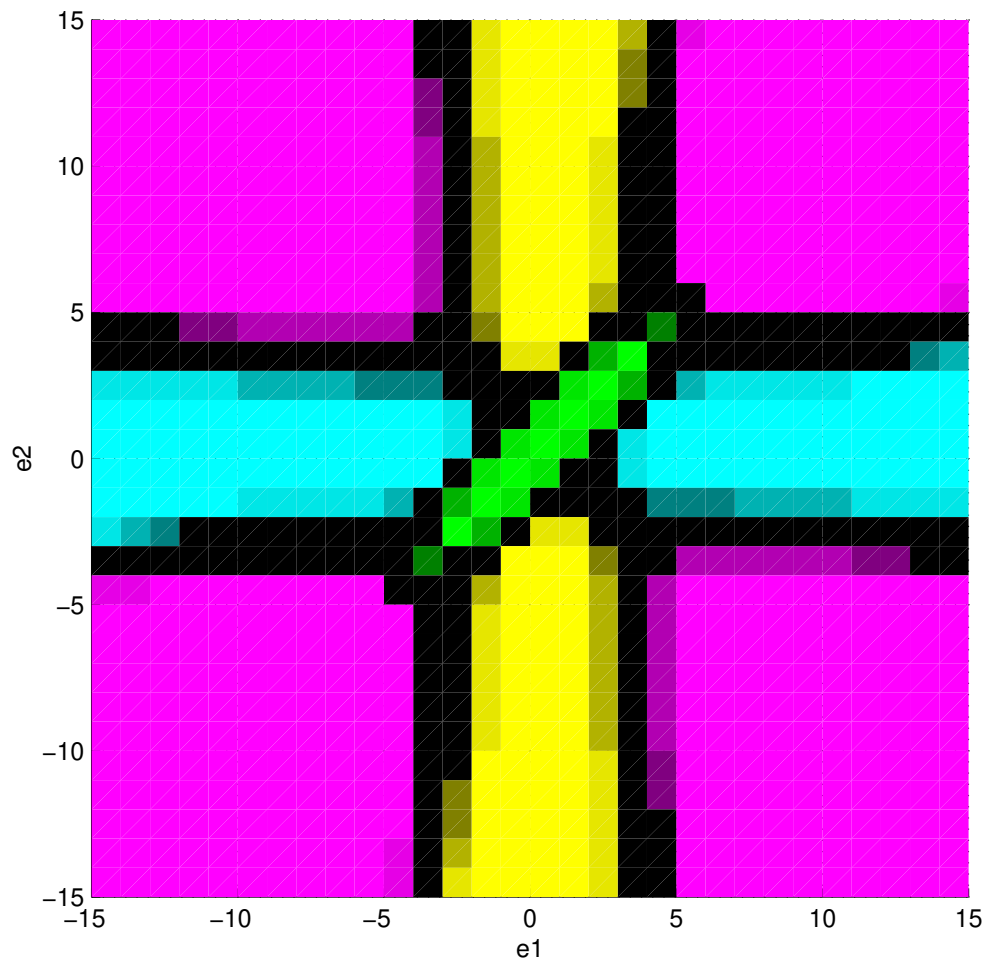


FIGURE G.3: *Posterior Probabilities, top view, showing classification boundaries (section 4.4.2). This figure is the combination of the four subfigures of figure G.2 viewed from directly above.*

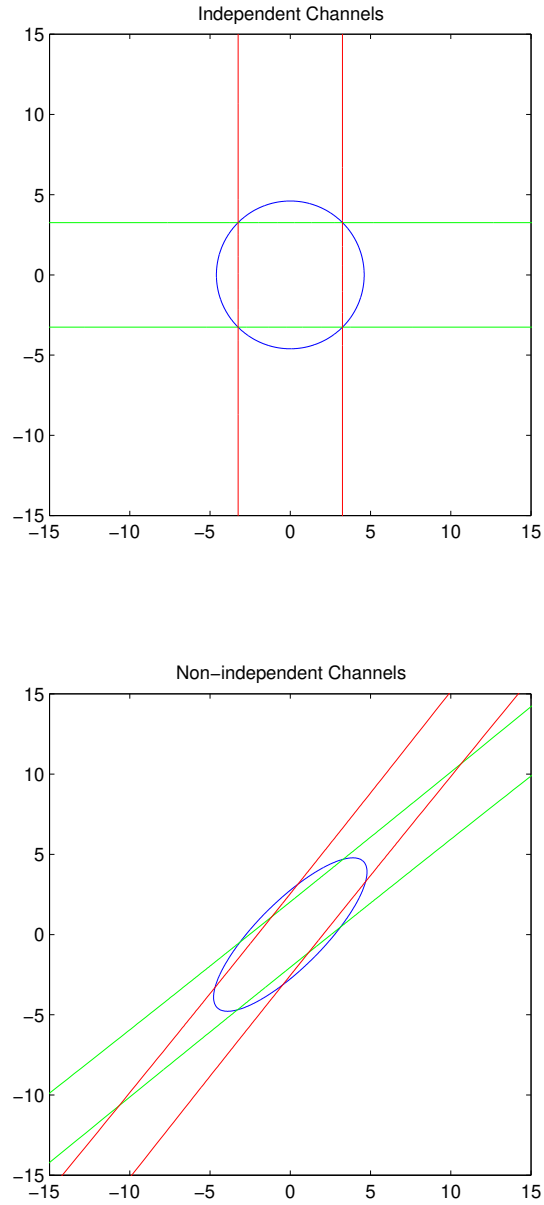


FIGURE G.4: Region $\mathcal{R}_{p_{11}}$ where both channels of a two-channel E-AR system are classified as uncorrupted (section 4.4.2). Region $\mathcal{R}_{p_{11}}$ is defined as the intersection of $\mathcal{R}_{p_{11} > p_{22}}$ (inside blue ellipse), $\mathcal{R}_{p_{11} > p_{21}}$ (between red lines) and $\mathcal{R}_{p_{11} > p_{12}}$ (between green lines). Two cases are shown; the first for independent channels, and the second illustrating the modification of the region for the non-independent two-channel detector. The lines in the lower figure which appear straight are in fact hyperbolae with distant foci.

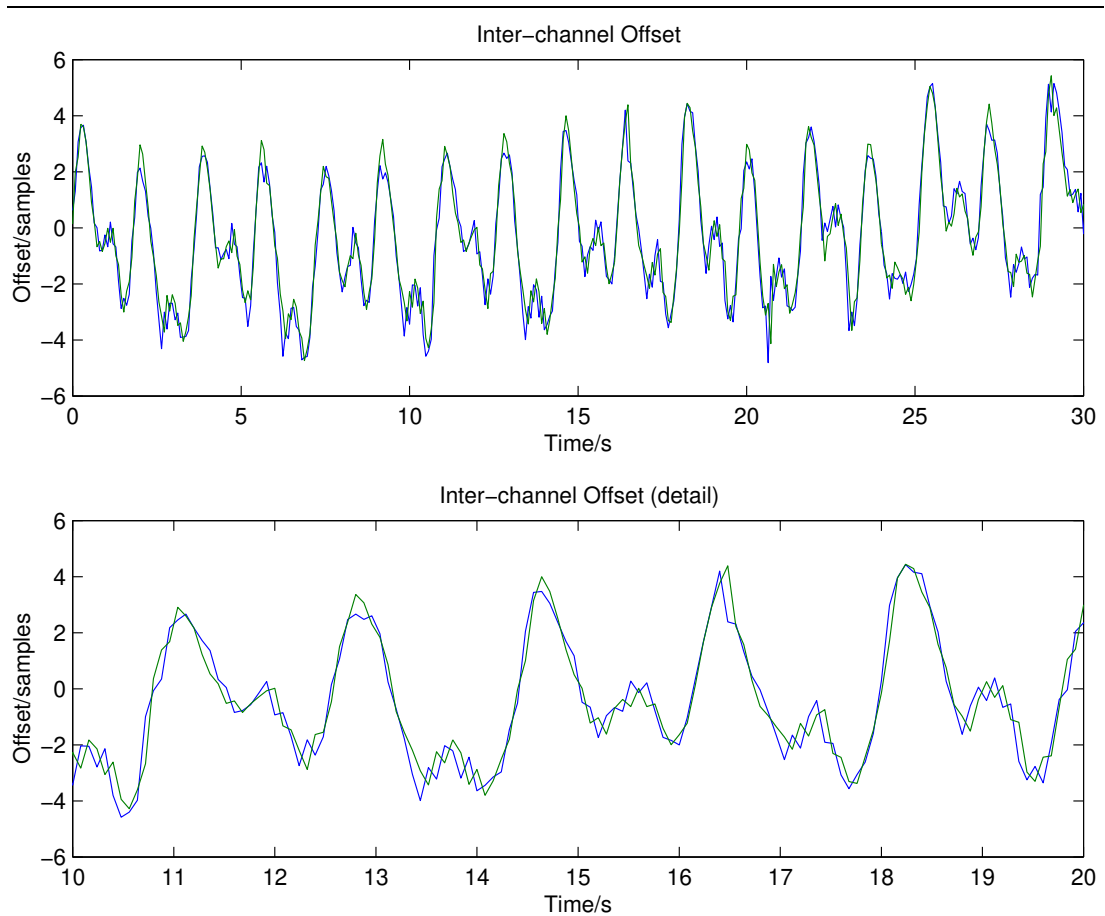


FIGURE G.5: $AR(1)$ plus sinusoidal basis (2 harmonics) as a model for inter-channel time offset (section 6.7). The blue line is the measured data, and the green predicted by the model.

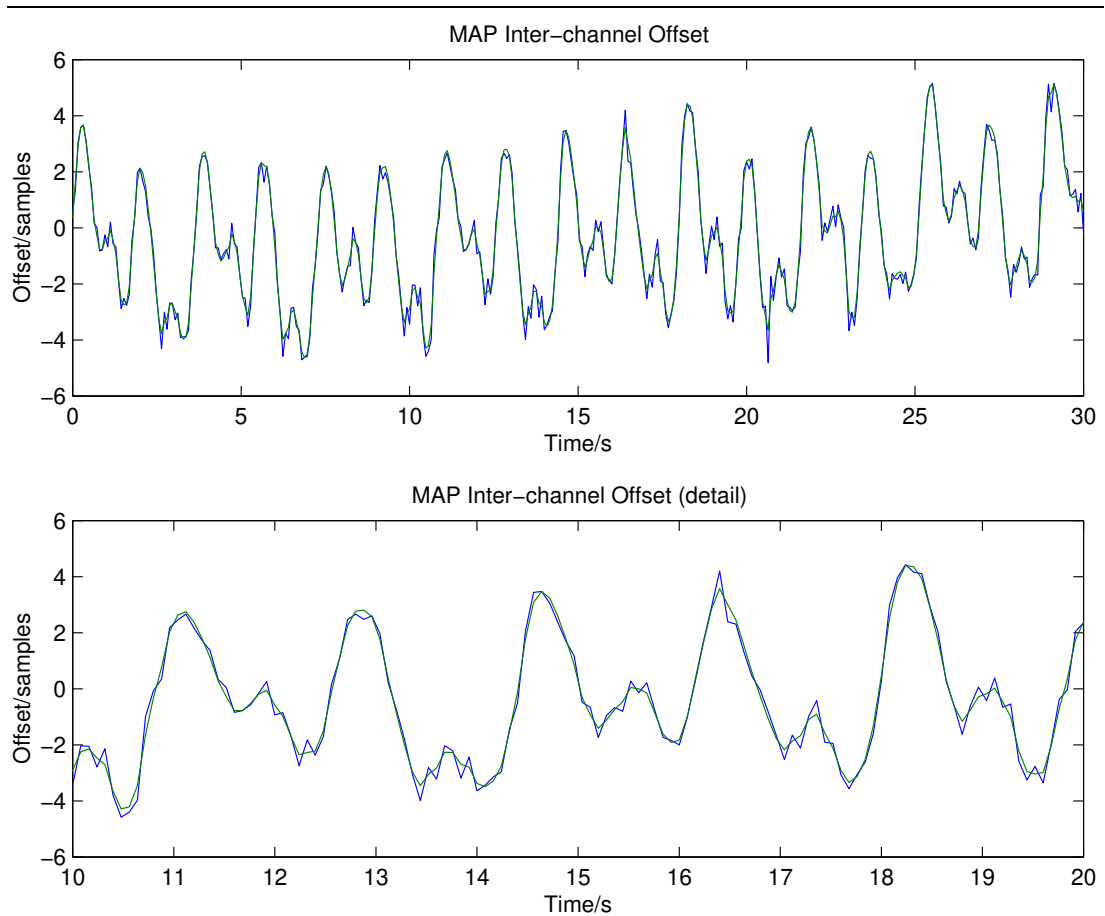


FIGURE G.6: MAP offset estimate from equation 6.65 with $AR(1)$ plus sinusoidal basis (2 harmonics) as a model for inter-channel time offset (section 6.9.1). The blue line is the raw measured offset, and the green is the MAP estimate based on the measured data and the model-based prior.

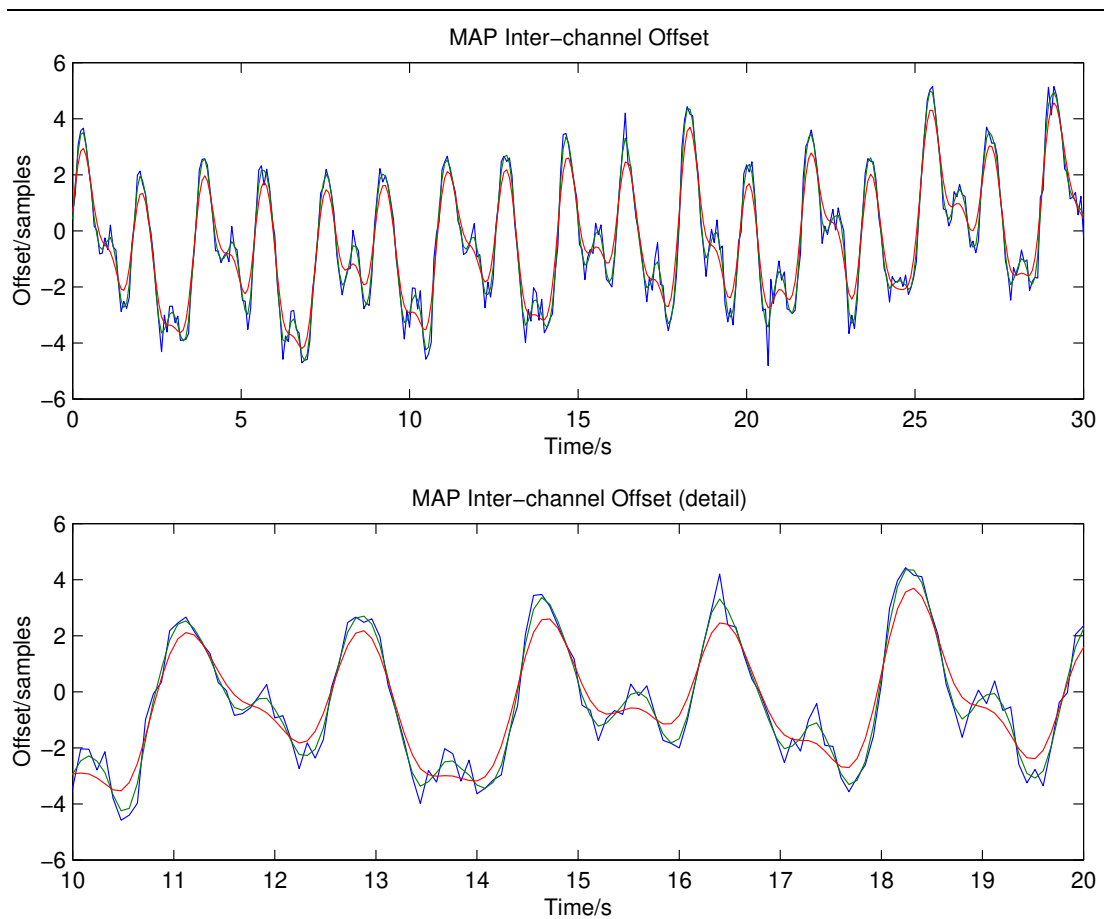


FIGURE G.7: MAP offset estimate using differential smoothness prior. The blue line is the raw measured offset. The green line is the MAP estimate based on the measured data and the differential smoothness prior with $\alpha = 1$, and the red line with $\alpha = 10$.

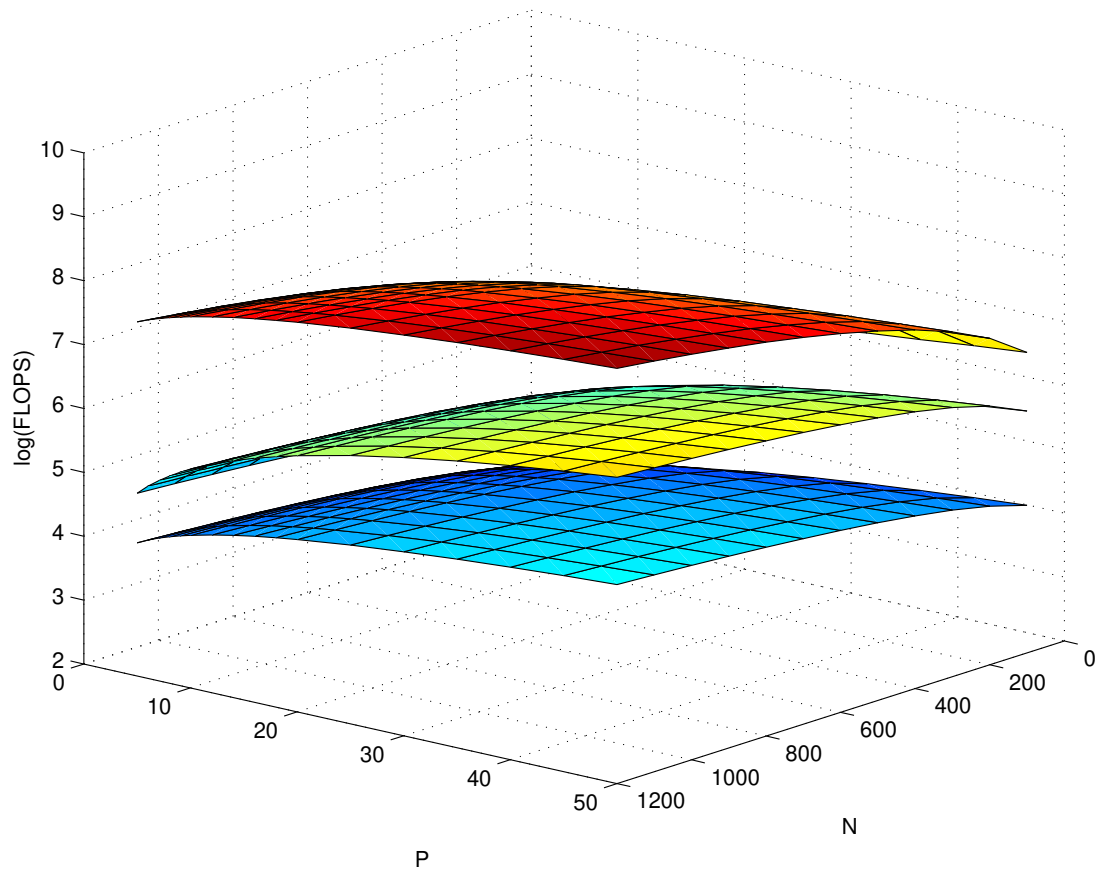


FIGURE G.8: Operations count (logarithm to base 10) for TLS (top), LS (middle) and Toeplitz LS (bottom) algorithms (section E.4). Over this range of problem sizes the TLS complexity is almost independent of the model order P as it is dominated by the data vector length N .

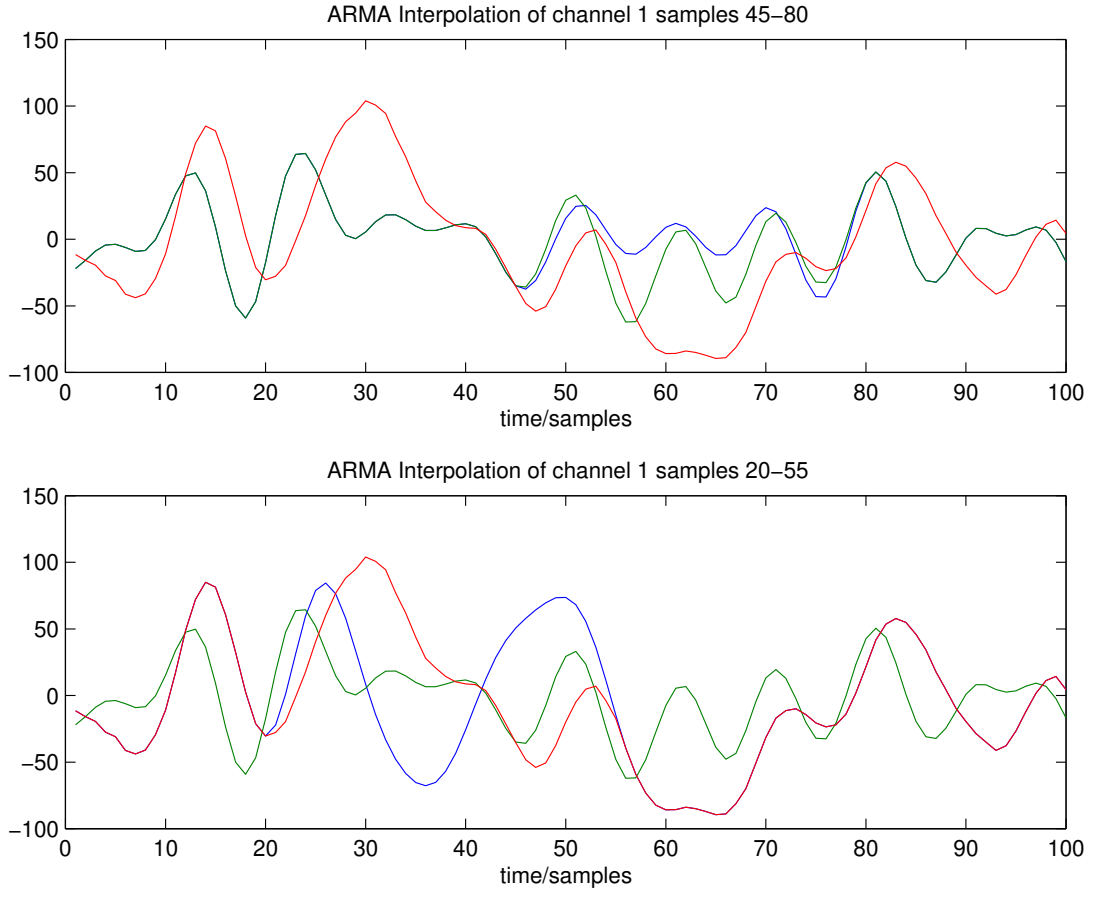


FIGURE G.9: *Independent ARMA interpolations of two-channel data (section 5.5). The green and red lines are channels 1 and 2 of a C-ARMA system. Sections of each have been interpolated (blue lines) using an independent ARMA model for each channel.*

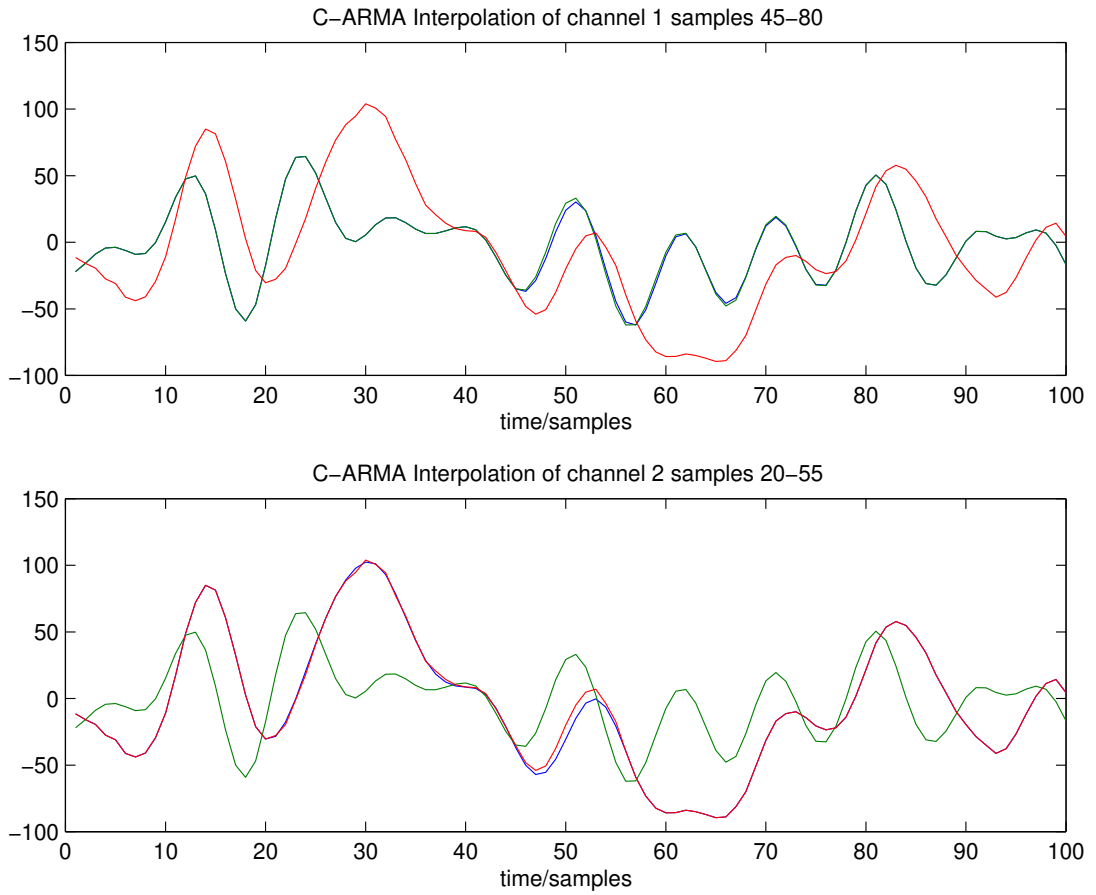


FIGURE G.10: Two-channel *C-ARMA* interpolation of synthetic data (section 5.5). The green and red lines are channels 1 and 2 of a *C-ARMA* system. Overlapping sections of each have been assumed unknown and interpolated (blue lines) using the joint *C-ARMA* interpolator.

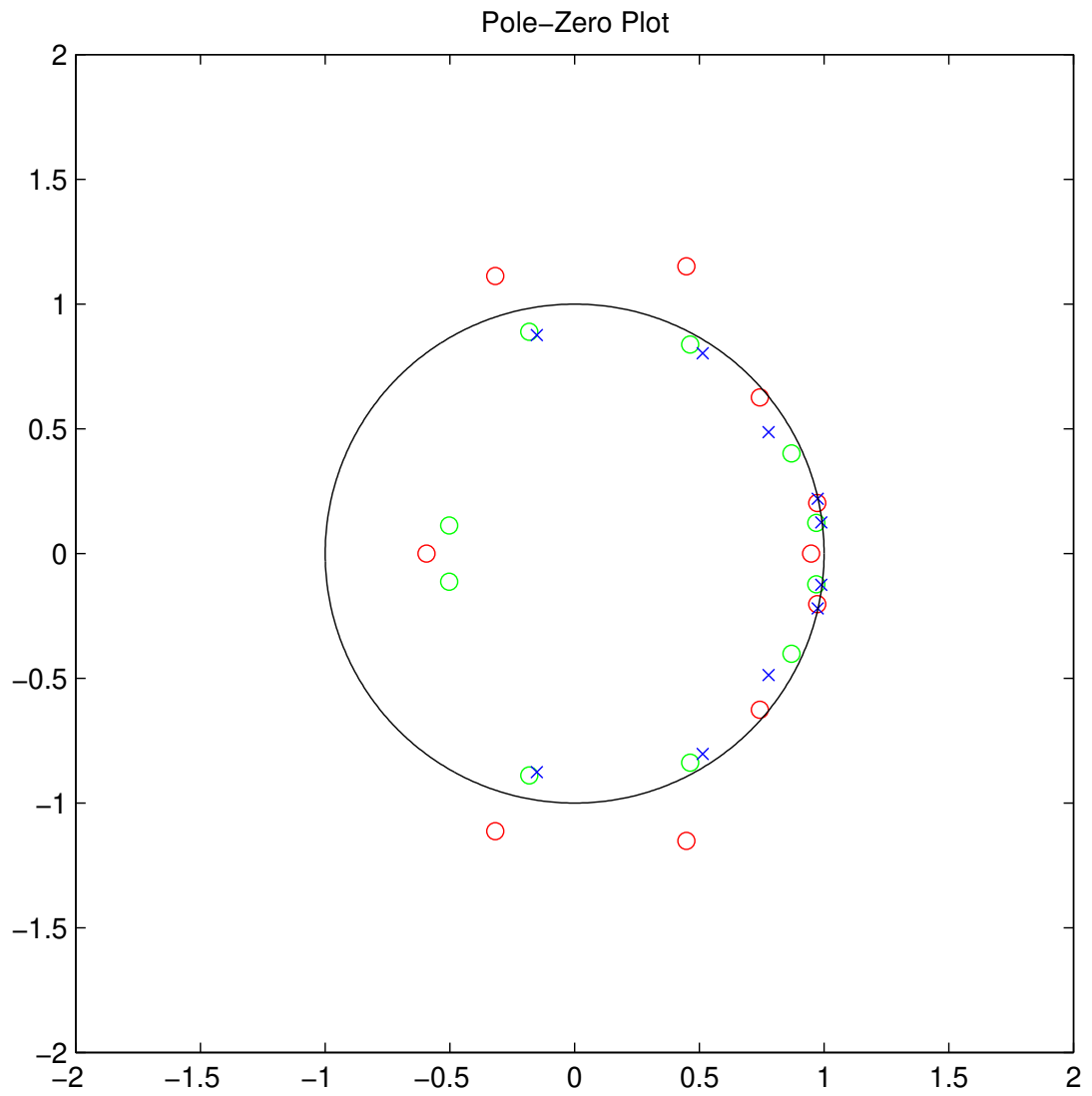


FIGURE G.11: *Estimated poles and zeros for C-ARMA model (section 5.5). The green and red circles show the positions of the estimated zeros for the left and right channel signals respectively. The model poles are shown by blue crosses.*

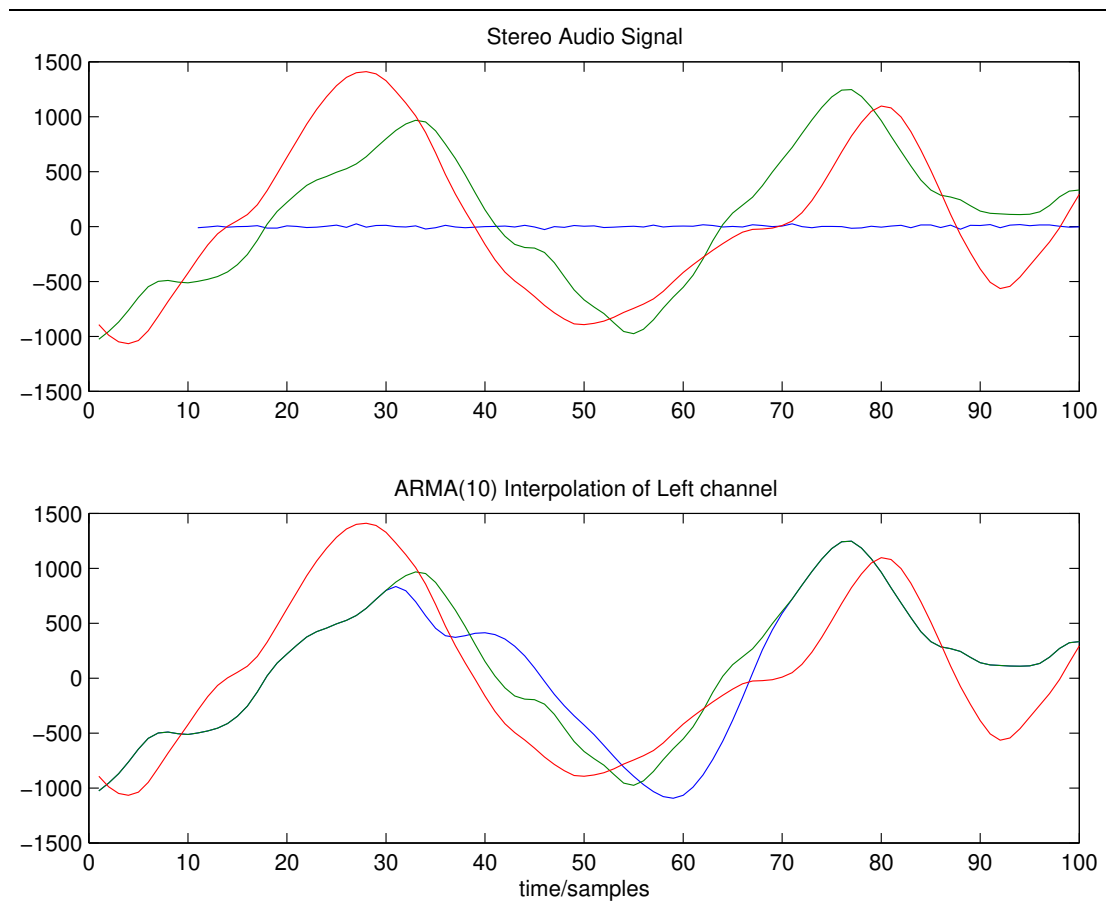


FIGURE G.12: ARMA interpolation of stereo audio data (section 5.5). The green and red lines are the left and right signals of a genuine stereo audio signal. The blue line in the upper figure shows the modelling error. The blue line in the lower figure is the order-10 ARMA interpolation of 40 samples of the left channel.

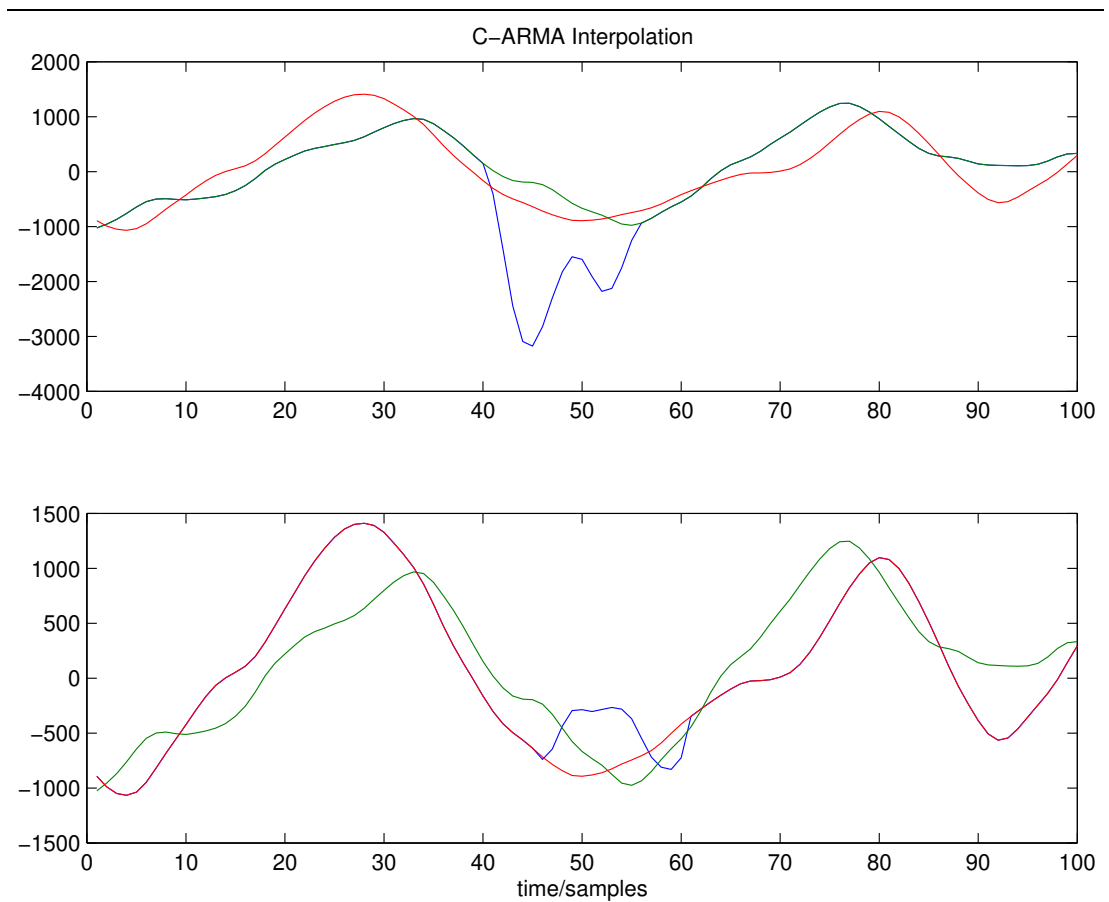


FIGURE G.13: *C-ARMA interpolation of stereo audio data (section 5.5). The green and red lines are the left and right signals of a genuine stereo audio signal. The blue lines show joint order-10 C-ARMA interpolations of the stereo data.*

References & Bibliography

- [1] ABBAGNARO, L. A., Ed. *Microphones: Collected Papers from the Journal of the Audio Engineering Society*. Audio Engineering Society, October 1979.
- [2] ADAMS, R., AND KWAN, T. VLSI architectures for asynchronous sample rate conversion. *Audio Engineering Society preprint no. 3355* (October 1992).
- [3] ADAMS, R., AND KWAN, T. Theory and VLSI implementation of asynchronous sample rate converters. *Audio Engineering Society preprint no. 3570* (March 1993).
- [4] ADRIEN, J. M., CAUSSE, R., AND DUCASSE, E. Sound synthesis by physical models—application to strings. *Audio Engineering Society preprint no. 2625* (March 1988).
- [5] AKAIKE, H. A new look at statistical model identification. *IEEE Transactions on Automatic Control* 19 (1974), 716–723.
- [6] AKAIKE, H. A Bayesian extension of the minimum AIC procedure of autoregressive modelling. *Biometrika* 66, 2 (February 1979), 237–242.
- [7] ARAKAWA, K., FENDER, D. H., HARASHIMA, H., MIYAKAWA, H., AND SAITOH, Y. Separation of a non-stationary component from the EEG by a nonlinear digital filter. *IEEE Transactions on Biomedical Engineering* (1986).
- [8] AXON, P. E., AND DAVIES, H. A study of frequency fluctuations in sound recording and reproducing systems. *Proceedings of the Institute of Radio Engineers* 96 (January 1949), 65–75.

- [9] BLESSER, B., LOCANTHI, B., AND STOCKHAM, T. G., Eds. *Digital Audio; collected papers from the AES Premier Conference*. Audio Engineering Society, June 1982.
- [10] BLUMLEIN, A. D. *British Patent 394,325: Improvements in and relating to sound-transmission, sound-recording and sound-reproducing systems*. In Ear-
gle [26], June 1933.
- [11] BOX, G. E. P., JENKINS, G. M., AND REINSEL, G. C. *Time Series Analysis, Forecasting and Control*, third ed. Prentice-Hall, 1994.
- [12] BRANDENBURG, K. *Perceptual Coding of High Quality Digital Audio*. In Brand-
denburg and Kahrs [14], 1998, p. 39.
- [13] BRANDENBURG, K., AND BOSI, M. Overview of MPEG audio: Current and
future standards for low-bit-rate audio coding. *Journal of the Audio Engi-
neering Society*, 1/2 (January/February 1997), 5–21.
- [14] BRANDENBURG, K., AND KAHR, M., Eds. *Applications of Digital Signal Pro-
cessing to Audio and Acoustics*. Kluwer Academic Publishers, 1998.
- [15] BRUEKERS, A. A. M. L., OOMEN, A. W. J., VAN DER VLEUTEN, R. J., AND
VAN DER KERKHOF, L. M. Lossless coding of 1-bit audio signals. *Eighth
regional convention of the Audio Engineering Society, Japan* (June 1997).
- [16] CANDY, J. V. *Signal Processing - The Model-based Approach*. McGraw-Hill,
1986.
- [17] CAPPÉ, O. Enhancement of musical signals degraded by background noise, using
long-term behaviour of the short-term spectral components. *Proceedings of
IEEE ICASSP* (1993).
- [18] CEDAR AUDIO LTD. *CEDAR for Windows*. 9 Clifton Court, Cambridge, CB1
7BN, UK, 1996.
- [19] CRAVEN, P., AND GERZON, M. Lossless coding for audio discs. *Journal of the
Audio Engineering Society* 44, 9 (October 1996), 706.
- [20] CROCHIERE, R. E., AND RABINER, L. R. *Multi-Rate Digital Signal Processing*.
Prentice-Hall, 1983.
- [21] CZYZEWSKI, A., KOSTEK, B., AND ZIELINSKI, S. New approach to the synthesis
of organ pipe sound. *Audio Engineering Society preprint no. 3957* (February
1995).
- [22] DE HOOG, F. A new algorithm for solving Toeplitz systems of equations. *Linear
Algebra and its Applications*, 88/89 (1987), 123–138.
- [23] DUDA, R. O., AND HART, P. E. *Pattern Classification and Scene Analysis*.
John Wiley and Sons, 1973.

- [24] DUNN, C., AND SANDLER, M. A simulated comparison of dithered and chaotic sigma-delta modulators. *Audio Engineering Society preprint no. 3926* (November 1994).
- [25] DURBIN, J. Efficient estimation of parameters in moving-average models. *Biometrika* 46 (1959), 306–316.
- [26] EARGLE, J., Ed. *Stereophonic Techniques: Collected Papers from the Journal of the Audio Engineering Society*. Audio Engineering Society, March 1986.
- [27] FAULKNER, T. A phased array. *HiFi News and Record Review* (July 1981).
- [28] FUCHS, H. Improving joint stereo audio coding by adaptive inter-channel prediction. *Proceedings of IEEE Workshop on Applications of Signal Processing in Audio and Acoustics* (1993).
- [29] GAYFORD, M., Ed. *Microphone Engineering Handbook*. Focal Press, 1994.
- [30] GERSHO, A. Principles of quantization. *IEEE Transactions on Circuits and Systems CAS-25* (July 1978), 427–436.
- [31] GERZON, M., CRAVEN, P., STUART, J., AND WILSON, R. Psychoacoustic noise shaped improvements in CD and other linear digital media. *Audio Engineering Society preprint number 3501* (March 1993).
- [32] GERZON, M., AND CRAVEN, P. G. Optimal noise shaping and dither of digital signals. *Audio Engineering Society preprint number 2822* (October 1989).
- [33] GERZON, M. A. The design of precisely coincident microphone arrays for stereo and surround sound. *Fiftieth convention of the Audio Engineering Society, preprint 20* (March 1975).
- [34] GILKS, W. R., RICHARDSON, S., AND SPIEGELHALTER, D. J. *Markov Chain Monte Carlo in practice*. Chapman and Hall, 1996.
- [35] GODSILL, S. J. *Restoration of Degraded Audio*. PhD thesis, University of Cambridge, 1992.
- [36] GODSILL, S. J. The restoration of pitch variation defects in gramophone recordings. *Proceedings of IEEE Workshop on Applications of Signal Processing in Audio and Acoustics* (1993).
- [37] GODSILL, S. J., AND RAYNER, P. J. W. *Chapter 6, Hiss Reduction*. In [39], 1998, pp. 135–149.
- [38] GODSILL, S. J., AND RAYNER, P. J. W. *Chapter 8, Restoration of Pitch Variation Defects*. In [39], 1998, pp. 171–190.
- [39] GODSILL, S. J., AND RAYNER, P. J. W. *Digital Audio Restoration*. Springer-Verlag, 1998.

- [40] GODSILL, S. J., AND RAYNER, P. J. W. *Section 5.4.2, ARMA model-based interpolation*. In [39], 1998, pp. 122–126.
- [41] GODSILL, S. J., AND RAYNER, P. J. W. Statistical reconstruction and analysis of autoregressive signals in impulsive noise using the Gibbs sampler. *IEEE Transactions on Speech and Audio Processing* 6, 4 (July 1998), 352–372.
- [42] GOLUB, G. H., AND VAN LOAN, C. F. *Matrix Computations*. The John Hopkins University Press, 1989.
- [43] GRAY, R. Quantization noise spectra. *IEEE Transactions on Information Theory IT-36* (November 1990), 1220–1244.
- [44] HAYKIN, S. *Adaptive Filter Theory*, second ed. Prentice-Hall, 1991.
- [45] HEIN, S., AND ZAKHOR, A. *Sigma Delta Modulators*. Kluwer Academic Publishers, 1993.
- [46] HICKS, C. M. The application of dither and noise-shaping to Nyquist rate digital audio. <http://www.eng.cam.ac.uk/~cmh/nspaper.html> (1995).
- [47] HICKS, C. M. Programmable DSP architectures. In *Audio—The second Century* (1999), Audio Engineering Society.
- [48] HICKS, C. M., AND GODSILL, S. J. A dual-channel approach to the removal of impulsive noise from archive recordings. *Proceedings of IEEE ICASSP* (1994).
- [49] HICKS, C. M., AND REID, G. The evolution of broadband noise reduction techniques. *Audio Engineering Society preprint no. 4307* (1996).
- [50] JECKLIN, J. A different way to record classical music. *Journal of the Audio Engineering Society* 29, 5 (May 1981), 329–332.
- [51] JEFFREYS, H. *Theory of Probability*. Oxford University Press, 1939.
- [52] JOHNSTON, J. D. Estimation of perceptual entropy using noise masking criteria. *Proceedings of IEEE ICASSP* (1988).
- [53] JOHNSTON, J. D. Transform coding of audio signals using perceptual noise criteria. *IEEE J. Selected Areas in Communications* 6, 2 (February 1988).
- [54] KAHRs, M. *Digital audio system architecture*. In Brandenburg and Kahrs [14], 1998, p. 195.
- [55] KENDALL, E. Private communication, February 1999.
- [56] KOKARAM, A. C. *Motion Picture Restoration*. PhD thesis, University of Cambridge, 1993.
- [57] KOKARAM, A. C., AND GODSILL, S. J. *A system for reconstruction of missing data in image sequences using sampled 3D AR models and MRF motion priors*, vol. 2 of *Computer Vision ECCV. Springer Lecture Notes in Computer Science*. April 1996, pp. 613–624.

- [58] KONDOZ, A. *Digital Speech; coding for low bit rate communication systems*. John Wiley & Sons, 1994.
- [59] KREYSZIG, E. *Advanced Engineering Mathematics*, sixth ed. Wiley, 1988.
- [60] KREYSZIG, E. *Appendix 3*, sixth ed. In [59], 1988, p. A73.
- [61] LEVINSON, N. The Wiener RMS error criterion in filter design and prediction. *Journal of Mathematics and Physics* 25, 4 (January 1947), 261–278.
- [62] LIM, J. S., Ed. *Speech Enhancement*. Prentice-Hall signal processing series. 1983.
- [63] LIM, J. S., AND OPPENHEIM, A. V. All-pole modelling of degraded speech. *IEEE Transactions on Acoustics, Speech and Signal Processing ASSP-26*, 3 (June 1978).
- [64] LIM, J. S., AND OPPENHEIM, A. V. Enhancement and bandwidth compression of noisy speech. *Proceedings of IEEE* 67, 12 (December 1979).
- [65] LINK, B. Physical modelling on a parallel DSP system. *Audio Engineering Society preprint no. 3394* (October 1992).
- [66] LIPSCHITZ, S. P., WANNAMAKER, R. A., AND VANDERKOOY, J. Quantization and dither: a theoretical survey. *Journal of the Audio Engineering Society* 40, 5 (April 1992), 355–375.
- [67] MAGRATH, A. J., AND SANDLER, M. B. Efficient linearization of sigma-delta modulators with digital domain dithering. *Audio Engineering Society preprint no. 4105* (October 1995).
- [68] MAHER, R. On the nature of granulation noise in uniform quantization systems. *Journal of the Audio Engineering Society* (January 1992), 12–20.
- [69] MAHER, R. A method for extrapolation of missing digital audio data. *Audio Engineering Society preprint number 3715* (October 1993).
- [70] MAKHOUL, J. Linear prediction: A tutorial review. *Proceedings of IEEE* 63 (1975), 561–580.
- [71] MERIDIAN AUDIO. *Meridian Lossless Packing*. Stonehill, Stukeley Meadows, Huntingdon, Cambs., PE18 6ED, United Kingdom, February 1999.
- [72] MOLLOY, E. *High Fidelity Sound Reproduction*. Newnes, 1958.
- [73] MOORE, B. C. J. *An Introduction to the Psychology of Hearing*. Academic Press, 1989.
- [74] NG, S. L. Estimation of corrupted samples in autoregressive moving average (ARMA) data sequences. Master's thesis, University of Cambridge Engineering Department, May 1997.
- [75] O RUANAIDH, J. J. K., AND FITZGERALD, W. J. Interpolation of missing samples for audio restoration. *Electronics Letters* 30 (April 1994).

- [76] O RUANAIDH, J. K., AND FITZGERALD, W. J. *Numerical Bayesian methods applied to Signal Processing*. Springer, 1996.
- [77] PRESS, W. H., FLANNERY, B. P., TEUKOLSKY, S. A., AND VETTERLING, W. T. *Numerical Recipes in C*. Cambridge University Press, 1988.
- [78] PRIESTLEY, M. B. *Chapter 5, Estimation in the Time Domain*. In [79], 1981, pp. 359–364.
- [79] PRIESTLEY, M. B. *Spectral Analysis and Time Series*. Academic Press, 1981.
- [80] RABINER, L. R. *Digital techniques for changing the sample rate of a signal*. In Blesser et al. [9], June 1982, pp. 79–89.
- [81] RAJAN, J. J. *Time Series Classification*. PhD thesis, University of Cambridge, 1994.
- [82] RAJAN, J. J., AND RAYNER, P. J. W. Bayesian model order selection for the KL-Transform and the SVD. *Proceedings of IEEE ICASSP (1994)*.
- [83] RAYNER, P. J. W., AND GODSILL, S. J. The detection and correction of artifacts in archived gramophone recordings. *Proceedings of IEEE Workshop on Applications of Signal Processing in Audio and Acoustics (1991)*.
- [84] RISSANEN, J. Modelling by shortest data description. *Automatica 14 (1978)*, 465–471.
- [85] RODET, X. Musical sound signal analysis/synthesis: sinusoidal and residual, and elementary waveform models. *IEEE UK Symposium on applications of Time-Frequency and Time-Scale methods (August 1997)*, 111–120.
- [86] ROSSUM, D. Constraint based audio interpolators. *Proceedings of IEEE Workshop on Applications of Signal Processing in Audio and Acoustics (1993)*.
- [87] SCHUCHMAN, L. Dither signals and their effect on quantization noise. *IEEE Transactions on Communications COM-12 (December 1964)*, 162–165.
- [88] SERRA, X. *A system for sound analysis/transformation/synthesis based on a deterministic plus stochastic decomposition*. PhD thesis, Stanford University, October 1989.
- [89] SHANNON, C. A mathematical theory of communication. *The Bell System Technical Journal 27 (July, October 1948)*, 379–423, 623–656.
- [90] SHANNON, C. Communication in the presence of noise. *Proceedings of the Institute of Radio Engineers 37, 1 (January 1949)*, 10–21.
- [91] SMITH, A. F. M., AND SPIEGELHALTER, D. J. Bayes factors and choice criteria for linear models. *Journal of the Royal Statistical Society 42 (1980)*, 213–220.
- [92] SPENCER, P. S. *System Identification with Application to the Restoration of Archived Gramophone Recordings*. PhD thesis, University of Cambridge, 1990.

- [93] STREICHER, R., AND DOOLEY, W. Basic stereo microphone perspectives: A review. *Journal of the Audio Engineering Society* 7/8 (July/August 1985), 548–556.
- [94] STUDIO AUDIO AND VIDEO LTD. *The SADiE Disk Editor*. The Old School, Stretham, Ely, UK, 1994.
- [95] TEWKSBURY, S., AND HALLOCK, R. Oversampled, linear predictive and noise-shaping coders of order N greater than 1. *IEEE Transactions on Circuits and Systems CAS-25*, 7 (July 1978), 436–447.
- [96] THEILE, G. Das Kugelflächenmikrofon. *Tonmeistertagung* (1986).
- [97] THERRIEN, C. W. *Discrete Random Signals and Statistical Signal Processing*. Prentice-Hall, 1992.
- [98] THERRIEN, C. W. *Section 9.5, ARMA Modelling*. In [97], 1992, pp. 550–575.
- [99] TROUGHTON, P., AND GODSILL, S. A reversible jump sampler for autoregressive time series. *Proceedings of IEEE-ICASSP* (1998), 2257–2260.
- [100] VALIÈRE, J. C. *La Restauration d'Enregistrements Anciens par Traitement Numérique—Contribution à l'étude de Quelques techniques récentes*. PhD thesis, Université du Maine, 1991.
- [101] VASEGHI, S. V. *Algorithms for Restoration of Archived Gramophone Recordings*. PhD thesis, University of Cambridge, 1988.
- [102] VELDHUIS, R. *Restoration of Lost Samples in Digital Signals*. Prentice-Hall, 1990.
- [103] WANNAMAKER, R. Psychoacoustically optimal noise shaping. *Journal of the Audio Engineering Society* 40, 7/8 (July/August 1992), 611–620.
- [104] WANNAMAKER, R., LIPSCHITZ, S., AND VANDERKOOY, J. Dithered quantizers with and without feedback. *Proceedings of IEEE Workshop on applications of Signal Processing to Audio and Acoustics* (October 1993).
- [105] WANNAMAKER, R., LIPSCHITZ, S., VANDERKOOY, J., AND WRIGHT, J. A theory of non-subtractive dither. *To appear in IEEE Transactions on Signal Processing*.
- [106] WIDROW, B., AND HOFF, M. E. Adaptive switching circuits. *IRE WESCON Convention* (1960), 96–104.
- [107] WILSON, P., AND WEBB, G. W. *Modern Gramophones and Electrical Reproducers*. Cassell, 1929.
- [108] ZÖLZER, U. Interpolation algorithms; theory and application. *Audio Engineering Society preprint no. 3898* (November 1994).