Helsinki University of Technology Laboratory of Acoustics and Audio Signal Processing

Espoo 2001 Report 61

FREQUENCY-WARPED AUTOREGRESSIVE MODELING AND FILTERING

Aki Härmä

TEKNILLINEN KORKEAKOULU
TEKNISKA HÖGSKOLAN
HELSINKI UNIVERSITY OF TECHNOLOGY
TECHNISCHE UNIVERSITÄT HELSINKI
UNIVERSITE DE TECHNOLOGIE D'HELSINKI

Helsinki University of Technology Laboratory of Acoustics and Audio Signal Processing
Espoo 2001 Report 61

FREQUENCY-WARPED AUTOREGRESSIVE MODELING AND FILTERING

Aki Härmä

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission for public examination and debate in Auditorium S4, Department of Electrical and Communications Engineering, Helsinki University of Technology, Espoo, Finland, on the 25th of May, 2001, at 12 o'clock noon.

Helsinki University of Technology
Department of Electrical and Communications Engineering
Laboratory of Acoustics and Audio Signal Processing

Teknillinen korkeakoulu
Sähkö- ja tietoliikennetekniikan osasto
Akustiikan ja äänenkäsittelytekniikan laboratorio

Helsinki University of Technology
Laboratory of Acoustics and Audio Signal Processing
P.O.Box 3000
FIN-02015 HUT
Tel. +358 9 4511
Fax +358 9 460 224
E-mail lea.soderman@hut.fi

ISBN 951-22-5457-3
ISSN 1456-6303

Otamedia Oy
Espoo, Finland 2001

HELSINKI UNIVERSITY OF TECHNOLOGY
P.O. BOX 1000, FIN-02015 HUT
http://www.hut.fi

ABSTRACT OF DOCTORAL DISSERTATION

Author Aki Sakari Härmä

Name of the dissertation
Frequency-warped autoregressive modeling and filtering

Date of manuscript May 14, 2001	Date of the dissertation May 25, 2001
□ Monograph	☑ Article dissertation (summary + original articles)

Department	Electrical and Communications Engineering
Laboratory	Laboratory of Acoustics and Audio Signal Processing
Field of research	Audio signal processing
Opponent(s)	Dr. Albertus den Brinker
Supervisor	Prof. Matti Karjalainen
(Instructor)	Docent Unto K. Laine

Abstract

This thesis consists of an introduction and nine articles. The articles are related to the application offrequency-warping techniques to audio signal processing, and in particular, predictive coding of wideband audio signals. The introduction reviews the literature and summarizes the results of the articles.

Frequency-warping, or simply warping techniques are based on a modification of a conventional signal processing system so that the inherent frequency representation in the system is changed. It is demonstrated that this may be done for basically all traditional signal processing algorithms. In audio applications it is beneficial to modify the system so that the new frequency representation is close to that of human hearing. One of the articles is a tutorial paper on the use of warping techniques in audio applications.Majority of the articles studies warped linear prediction, WLP, and its use in wideband audio coding. It is proposed that warped linear prediction would be particularly attractive method for low-delay wideband audio coding.

Warping techniques are also applied to various modifications of classical linear predictive coding techniques. This was made possible partly by the introduction of a class of new implementation techniques for recursive filters in one of the articles. The proposed implementation algorithm for recursive filters having delay-free loops is a generic technique. This inspired to write an article which introduces a generalized warped linear predictive coding scheme. One example of the generalized approach is a linear predictive algorithm using almost logarithmic frequency representation.

Keywords warping, delay-free loops, linear prediction, WLP, LPC, warped filters, audio signal processing

UDC 004.383.3:621.391:534.7	Number of pages 149
ISBN (printed) 951-22-5457-3	ISBN (pdf) 951-22-5460-3
ISBN (others)	ISSN 1456-6303

Publisher Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing

Print distribution

☑ The dissertation can be read at http://www.hut.fi/Yksikot/Kirjasto/Diss/

Preface

This thesis is a collection of nine articles. One of them was my second own publication written in the fall of 1996 and the most recent one was submitted in January 2000. This thesis is my *collected papers* covering the years 1997 to 2001. Most researchers have a need to recycle their previous works and I am not completely free of that. Therefore, I would like to express my deepest gratitude to Helsinki University of Technology, and Laboratory of Acoustics and Audio Signal Processing for this unique opportunity to spread out my old papers and equip them with upgraded explanations. They may even give me a grade for that.

The work reported in this thesis was not done alone or in isolation. Unto K. Laine introduced me to this field and has been leading and guiding my work for many years. His role as a teacher and partner in research work has been of fundamental importance. Matti Karjalainen is the supervisor of this work and has also been a major contributor. Laine and Karjalainen appear as co-authors and co-innovators in many of the articles in this Thesis. Marko Juntunen introduced me to many new techniques and I am also grateful to him for critical reading of an early version of this work. The list of co-authors in my JAES article is impressive: five doctors from HUT/Acoustics. I am grateful to all of them. Having them backing up my paper, the review process was a piece of cake. Paavo Alku and Vesa Välimäki have also been very helpful and performed critical reading of many of my manuscripts. I would like to thank also many other people at HUT/Acoustics I've had a privilege to work with in the recent years. Matti Karjalainen and Lea Söderman deserve special thanks for all last-minute arrangements regarding this thesis. My current colleagues at Media Signal Processing Research, Agere Systems (which was still a part of Bell Labs in February) have also been very kind and supportive. In particular, I am thankful to my supervisor Peter Kroon.

Ioan Tabus from Tampere University of Technology, and Keiichi Tokuda from Nagoya Institute of Technology performed the pre-examination of this work. I highly respect their experience and knowledge of the field. They made many good remarks and suggestions and I am really glad that I got their acceptance for printing this thesis.

I would like to thank my lovely wife, Laura, and our children Mandi, Juri, and Jalo for their support and patience. Jalo was five days old at the time of writing this preface. My parents and sisters also deserve thanks for all support, and especially for their help in organizing 'karonkka'.

This work has been supported by the Academy of Finland and graduate school GETA. Additional support have been provided by *Jenny ja Antti Wihurin säätiö*, *Nokia Foundation*, *Instrumentariumin tiedesäätiö*, and *Tekniikan edistämissäätiö*. *Comsol Finland* is also acknowledged for valuable software support.

Aki Härmä, Springfield, New Jersey, USA

List of Symbols

a_k, b_k, c_k	kth coefficient
$e(n)$	residual (prediction error signal)
i	Imaginary unit $\sqrt{-1}$
$s_d(n)$	deterministic signal
$s_r(n)$	regular signal
$r(n)$	fundamental sequence
$x(n)$	audio or speech signal
$\tilde{x}(n)$	estimate for a signal (prediction)
$A(z)$	Z-transfom of a FIR filter
$C_{k,p}$	element in a covariance matrix
$D_k(z)$	Z-transform of a kth subfilter
$E(z)$	Z-transform of a prediction error signal, or residual
$E_q(z)$	Z-transform of residual quantization error signal
K_p	pth reflection coefficient of a lattice filter
L	the order of a filter
$P(z)$	Z-transform of a predictor
$P(\omega)$	The power spectrum of a signal
R_k	kth autocorrelation term
$X(z)$	Z-transform of a signal
$X_q(z)$	Z-transform of a coding error signal
$Q[\cdot]$	Quantization operator
$E[\cdot]$	Expectation operator
γ	the parameter used in bandwidth widening
ω	Angular frequency
$\phi_\ell(n)$	ℓth basis function in TVAR models.
$\psi(\omega)$	frequency-warping function
λ	the warping parameter

List of abbreviations

AP	All-pass filter
AR	Autoregressive
ARMA	Autoregressive moving average
CELP	Code Excited Linear Prediction
DSP	Digital Signal Processing
ERB	Equivalent Rectangular Bandwidth
GAL	Gradient Adaptive Lattice
HRTF	Head Related Transfer Function
MPEG	Moving Pictures Expert Group
FAM	Frequency-Amplitude Modulated complex exponential
FFT	Fast Fourier Transform (Cooley & Tukey 1965)
FIR	Finite Impulse Response
IIR	Infinite Impulse Response
ISO	International Standardization Organization
ITU	International Telecommunication Union
LAR	Logarithmic Area Ratio
LMS	Least mean square
LSF	Line Spectral Frequency
LSP	Line Spectrum Pair
LP	Linear Prediction
LPC	Linear Predictive Coding
LS	Least Square
MA	Moving Average
MMSE	Minimum Mean Square Error
PCM	Pulse Code Modulation
WFIR	Warped FIR-type filter
WIIR	Warped IIR-type filter
WLP	Warped Linear Prediction
WLPC	Warped Linear Predictive Coding

List of publications

This thesis summarizes the following articles and publications and ties them to earlier contributions on the field. In text, they are referred to as [P1]–[P9].

[P1] A. Härmä, U. K. Laine, and M. Karjalainen, “An experimental audio codec based on warped linear prediction of complex valued signals,” in *Proc IEEE Int. Conf. Acoust. Speech and Signal Proc. (ICASSP’99)*, vol. 1, (Munich), pp. 323–327, 1997.

[P2] A. Härmä, U. K. Laine, and M. Karjalainen, “Backward adaptive warped lattice for wideband stereo coding,” in *Signal Processing IX: Theories and applications (EUSIPCO’98)*, (Greece), pp. 729–732, 1998.

[P3] A. Härmä, U. K. Laine and M. Karjalainen, “On the utilization of overshoot effects in low-delay audio coding,” in *Proc. IEEE Int. Conf. Acoust. Speech and Signal Proc. (ICASSP’99)*, vol. II, (Phoenix, Arizona), pp. 893–896, IEEE, March 1999.

[P4] A. Härmä and U. K. Laine, “Warped low-delay CELP for wide-band audio coding,” in *Proc. of the AES 17th Int. Conference: High-Quality Audio Coding*, (Florence, Italy), pp. 207–215, September 2-5 1999.

[P5] A. Härmä, M. Juntunen, P. Kaipio, “Time-varying autoregressive modeling of audio and speech signals”, *Signal Processing X: theories and applications (EUSIPCO 2000)*, (Tampere, Finland), pp. 2037-2040, September 2000.

[P6] A. Härmä, “Implementation of frequency-warped recursive filters,” *Signal Processing*, 80 (3), pp. 543-548, February 2000.

[P7] A. Härmä, “Linear predictive coding with modified filter structures,” Report no. 59/Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, Espoo, Finland, 2001.(*Submitted to IEEE Trans. Speech and Audio Processing*, January 2000.)

[P8] A. Härmä and U .K. Laine, “A Comparison of warped and conventional linear predictive coding”, *IEEE Trans. Speech and Audio Processing*, July 2001.

[P9] A. Härmä, M. Karjalainen, L. Savioja, V. Välimäki, U. K. Laine, and J. Huopaniemi, “Frequency-warped signal processing for audio applications”, J. Audio Eng. Soc., Vol. 48 (11), pp. 1011-1031, November 2000.

Table of Contents

Preface i

List of symbols ii

List of abbreviations iii

List of publications iv

1 Introduction **1**

1.1 Scope of this thesis 1

1.2 Coding of audio and speech signals 2

1.2.1 Attributes and applications 2

1.2.2 Techniques for coding 4

1.3 Contents of this thesis 7

2 Theoretical background **8**

2.1 Linear stationary signal models 8

2.1.1 Wold decomposition theorem 9

2.1.2 Prediction problem 9

2.2 Linear prediction 10

2.2.1 Linear predictive coding 13

2.2.2 Spectral representation 14

2.2.3 Lattice methods 16

2.3 Linear nonstationary signal models 16

2.3.1 Frame-based processing 17

2.3.2 Adaptive filtering 18

2.3.3 Deterministic regression time-varying LPC 19

3 Contributions of this thesis **20**

3.1 Generalized predictor structures 20

3.2 Frequency-warped signal processing 23

3.3 Warped linear prediction 24

3.4 Low-delay audio coding 26

3.5 Time-varying spectral modeling 28

3.6 Future work 28

4 Conclusions **30**

4.1 Main results of this thesis 30

4.2 Contribution of the author 30

5 Publications **41**

1. Introduction

1.1 Scope of this thesis

Encoding is a process of changing the representation of a signal for transmission or storage so that it meets the requirements of the media. The objective in *decoding* is to reconstruct the original signal from this representation so that the quality of the signal, in respect to some measure, is not deteriorated. *Codec*[1] is a common noun for an encoder-decoder system.

The techniques presented in this thesis have been developed for *lossy* coding of speech and audio signals where the deterioration of a reproduced signal is ultimately judged by the human ear. Due to the limitations of the hearing mechanism, see, e.g., (Moore 1997), a *technically* deteriorated signal can be perceived as faultless[2]. A *perceptual codec* is a lossy encoder-decoder system which is designed so that it utilizes the properties of human perception (Jayant, Johnston & Sefranek 1993). *Lossless* coding, see, e.g., (Gerzon, Graven, Stuart, Law & Wilson 1999), where a decoder can reproduce the original signal exactly, is not in the scope of this thesis. The main requirement of the media is that the bitrate of an encoded bit-stream should be lower than that of the original signal. Error concealment and channel coding methods (Lin & Costello Jr. 1983) for noisy transmission or storage media are not studied in this thesis.

Technically, this thesis concentrates to the class of *frequency-warped* digital signal processing, DSP, techniques (Oppenheim, Johnson & Steiglitz 1971, Strube 1980). This is a relatively generic framework which can be applied to many conventional DSP techniques to produce new tools where some aspects of human hearing can be automatically incorporated into the system. Even if this thesis principally addresses perceptual coding of audio and speech signals, an introduction to other potential applications of this methodology is also given [P9].

[1] Or coder.

[2] Some authors call this *perceptually lossless* coding (Scheirer 1999)

1.2 Coding of audio and speech signals

Classically, the fields of audio and speech coding have been somewhat different. This is because speech codecs utilize speech-specific features while audio codecs cannot generally rely on the characteristics of the input signal. In this thesis, the diversity of the field is not so pronounced because the presented methods are mainly based on conventional techniques in speech coding, but they are applied largely to wideband audio signals. On the other hand, most of the presented techniques are also directly applicable to speech coding systems. Although no speech-specific techniques, such as pitch prediction (Atal 1982), or voiced-unvoiced coding[3] (Atal & Hanauer 1971) are studied, the concept of audio coding is assumed to cover also many speech coding algorithms and applications. In fact, the diversity of the field of audio and speech coding techniques is not decreasing, in particular, modern low bit-rate audio and speech codecs are increasingly based on highly signal-dependent features, see, e.g., (Nishiguchi 1999, Scheirer 1999).

1.2.1 Attributes and applications

There is a large number of different applications for audio and speech coding. Transmission and storage of audio data are the two principal application types. The main attributes for an audio or speech codec are: bitrate, quality, delay, computational complexity and memory requirements, and processability. These are briefly discussed in the following subsections.

Bitrate

Reduction of bitrate is the primary motivation for the use of codecs[4]. In traditional communications applications it is usually necessary to maintain a constant bitrate while in storage and other non-real-time applications bitrate could be time-varying and depend on properties of the input signal, or requirements of the media.

In wideband 16-bit stereophonic audio at the sampling rate of 48 kHz, the bitrate is almost 1.6 Mbit/s. Current state-of-the-art wideband audio codecs can reduce this to 128 kbits/s so that the subjective quality is practically unaltered (Soulodre, Grusec, Lavoie & Thibault 1998). The MPEG-4 General Audio codec (Grill 1999), which is actually a large collection of different codecs, is capable to produce bitrates ranging from 2 to 64 kbits/s. Johnston (1988) estimated that sufficient bitrate for wideband audio would be around 2 bits per sample. In highly signal-dependent codecs the bitrate can be extremely low. For example, an ultimate speech codec would consist of a speech recognition system as encoder and a speech synthesizer as a decoder. Similar approach has also been proposed for compression of music signals, e.g., in (Scheirer 1999).

[3] or *vocoding*.

[4] In the AES 17th Conference (High-Quality Audio Coding) in Florence, a common way to start a conversation among 'the developers of audio coding techniques from the periphery of Europe' was: *Nice to meet you. What's your bitrate?'*.

Quality

When human listeners in good listening conditions cannot find a distinction between original test signals and outputs of a codec, the codec is called *transparent*(Johnston & Brandenburg 1992). Transparency is an important concept in the sense that it is a subjectively fixed attribute and makes comparison of different codecs straightforward[5]. It is significantly more challenging to evaluate different impairments in reproduced audio signals (Soulodre & Lavoie 1999). In practice, transparency is required only in very specific applications – those who need truly transparent coding can usually afford to transmit and store uncoded audio bitstreams[6] or use lossless coding techniques (Gerzon et al. 1999).

The human ear can perceive frequencies up to 20 kHz (Fletcher 1953). Muraoka, Iwahara & Yamada (1981) studied perception of reduction in bandwidth of musical signals. They found that most people hear the difference between full audio band and 14 kHz band but only few can hear the difference if the bandwidth is restricted to 18 kHz. In modern audio systems, typical audio bandwidths are above 20 kHz. It is usually assumed that sufficient bandwidth for speech is 10 kHz [7]. Naturally, the speech production system is capable of producing audible frequency components above this limit, e.g., in plosive sounds. The relation between subjective quality and bandwidth is highly nonlinear. For example, it has been demonstrated that perceived sound quality in a wideband codec may be higher than in a corresponding narrowband codec even if the latter would produce less audible distortion (Roy & Kabal 1991).

Delay

In most codecs it is necessary to use *buffering* which delays the processing of an input signal. This yields *algorithmic coding delay* which is an important attribute in many real-time applications. This topic was discussed in [P4], where it was estimated that a sufficiently low coding delay for most of applications would be around 2-10 milliseconds. Typically, the coding delay in wideband audio codecs ranges from 20 to 200 milliseconds. In low bit rate codecs the main source of algorithmic delay is related to *bit reservoir* techniques, where more bits are allocated to *difficult* parts of the input signal, while, e.g., pauses in music, can be coded with fewer bits.

Computational complexity and memory requirements

Although processors are becoming increasingly powerful and memory is getting cheaper and faster these are still fundamental requirements in most of practical applications for coding techniques. *Nathan's First Law*[8], which states that *"Software is a gas – it expands to fit the container it is in"*, applies to coding algorithms, too.

[5] Also in psychoacoustic experiments the goal is usually to find *just noticeable difference* (Zwicker & Fastl 1990, Moore 1997).

[6] For example, the Finnish Broadcasting Company, YLE, is currently converting their huge archive of recordings to an uncoded digital form. In addition, they don't use codecs in their production chain.

[7] Kleijn & Paliwal (1995*a*) cited to Denes & Pinson (1963), as a source of this information.

[8] By Nathan Myhrvold, Microsoft's former chief technology officer.

Processability

It is often necessary to apply various types of post-processing techniques to coded signals. In some applications it may be necessary to edit and combine coded bitstreams. This can be done with decoded signals but it would be desirable to do this directly with encoded material. Otherwise, the produced new material suffers from *tandem* coding artifacts. Recently, it have been demonstrated that it is possible to add new powerful functionalities, such as *pitch shifting* and time-scale modifications, to decoders where the parametrization of the signal is at high conceptual level, see, e.g., (Levine & Smith 1999).

1.2.2 Techniques for coding

There are many alternative taxonomies for different audio and speech coding techniques. In most of the available techniques, the emphasis in the coding process is to transmit spectral information[9]. Techniques for spectral analysis, see e.g., (Stoica & Moses 1997), are conventionally divided to two broad approaches: *non-parametric* and *parametric* techniques. Although many current codecs use partially techniques from both main branches, this is an illustrative way to classify different coding algorithms.

Non-parametric codecs

A typical example of a non-parametric codec is a subband or transform codec of Fig. 1.1. Here, a signal is first decomposed to spectral components using a filterbank or a transform. Each spectral component is quantized under the control of a *psychoacoustic model* which determines the *frequency masking* characteristics within each subband. This model allocates a different number of bits to each of the frequency bands. This scheme was introduced for speech coding by (Zelinski & Noll 1977), where they used Fast Fourier Transform, FFT, (Cooley & Tukey 1965) for spectral decomposition. Subband coding of speech signals had already been studied in (Crochiere, Webber & Flanagan 1976). Brandenburg, Langenbucher, Schramm & Seitzer (1982) applied this to wideband audio signals. A large number of different techniques have been proposed for subband decomposition, see, e.g., (Johnston & Brandenburg 1992, Brandenburg 1998), for review. Subband techniques based on *wavelet* transform have been used, e.g. in (Purat & Noll 1996, Hamdy, Ali & Tewfik 1996). Many wideband audio coding algorithms are also commercially available, e.g., AC-3[10] (Fielder, Bosi, Davidson, Davis, Todd & Vernon 1996), PAC (Johnston, Sinha, Dorward & Quackenbush 1996), ATRAC (Tsutsui, Suzuki, Shimoyoshi, Sonohara, Akagiri & Heddle 1996), and international ISO/IEC standards MPEG-1 (Brandenburg 1994, ISO/IEC 1993), MPEG-2 (Stoll 1996), and MPEG-4 (Grill 1999).

[9] However, there are deviants from this general line such as *waveform interpolation* codecs for speech (Kleijn & Paliwal 1995*b*) and scalar waveform quantization techniques (Moorer 1979).

[10] Currently, a part of Dolby Digital (Vernon 1999).

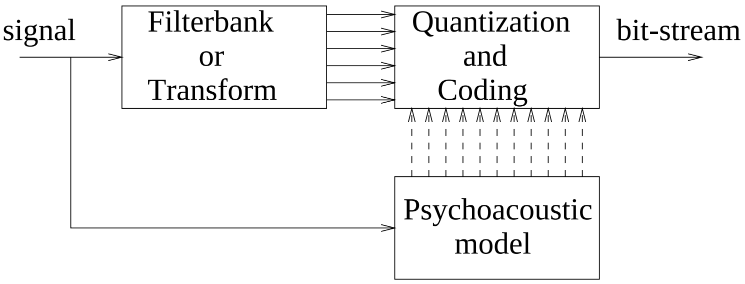


Figure 1.1: A typical subband or transform encoder based on subband decomposition and quantization controlled by a psychoacoustic model.

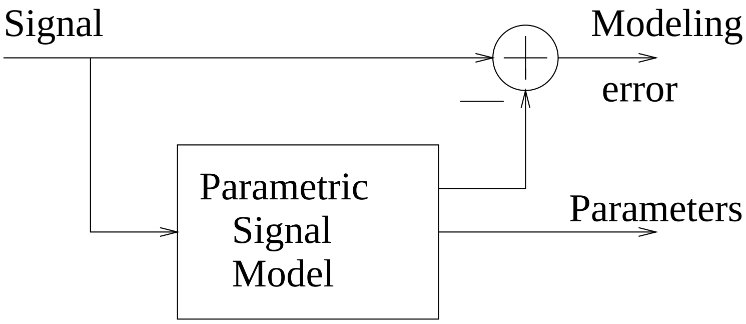


Figure 1.2: A parametric encoder.

Parametric codecs

Parametric coding is based on signal models. This approach is illustrated in Fig. 1.2. Here, the coding process involves estimation and coding of the parameters of the model. Often, it is also necessary to transmit the part of the signal which cannot be modeled, that is, the modeling error, as side information. There are many different variants for this scheme.

In classical *linear predictive coding*, LPC, the signal model is usually an allpole filter and an excitation signal, which may be interpreted as a modeling error signal. In some formulations of low-delay coding, the parameters need not to be transmitted but they can be estimated from the decoded signal. Most of the work in this thesis is related to LPC techniques. Linear prediction is a standard technique in speech codecs (Kleijn & Paliwal 1995*a*). In recent two decades especially Code Excited Linear Prediction, CELP (Schroeder & Atal 1985), and its many alternative formulations such as MELP (McCree & Barnwell III 1995), have been widely used in speech coding. In few cases, LPC has been used also for wideband audio coding, e.g., in (Singhal 1990, Boland & Deriche 1995).

There are many *hybrid* techniques which use both subband and LPC techniques. In TwinVQ [11] (Iwakami & Moriya 1996, Moriya, Iwakami, Ikeda & Miki 1996) and Transform Coded Excitation, TCX (Lefebvre, Salami, Laflamme & Adoul 1993, Bessette, Salami, Laflamme & Lefebvre 1999), linear predictive techniques are applied to spectral

[11] which has also been adopted to MPEG-4 general audio codec (Grill 1999).

parametrization but quantization is performed for a spectral representation of the remaining residual signal[12]. Multi-band excitation methods (Hardwick & Lim 1988)ăcan also be seen as a version of TCX. On the other hand, techniques where subband decomposition is followed by LPC applied separately to each subband have become a popular extension for audio coding algorithms (Lin & Steele 1993, Dimino & Morpurgo 1996).

Hedelin (1981) proposed a speech codec based on sinusoidal modeling where only dominant spectral peaks are coded as sets of parameters representing amplitude, frequency, and phase. This work was extended in (McAulay & Quatieri 1986) and applied to wideband audio signals by Smith & Serra (1987). In recent years, coding techniques where signal is decomposed into sinusoids, noise, and transients have been studied extensively, see, e.g., in (Hamdy et al. 1996, Purnhagen, Edler & Ferekidis 1998, Verma 1999). These techniques are particularly attractive for very low bit rate coding and they usually provide direct means for various additional functionalities, such as pitch shifting and time-scale modifications. However, this type of parametrization often requires long signal buffers, that is, the coding delay is high.

Comparison or parametric and non-parametric approaches

According to the information theory, see, e.g., (Berger & Gibson 1998), a parametric representation for a signal is more efficient than a *blind* non-parametric representation if the parameters are those of an appropriate source model for a signal. For example, a linear predictive model assumes that the signal is an autoregressive process, i.e., a white noise signal filtered by a finite-order allpole filter[13]. In speech coding, the success of LPC have been explained by the fact that an allpole model is a reasonable approximation for the transfer function of the vocal tract (Atal & Hanauer 1971). Allpole model is also appropriate in terms of human hearing because the ear is more sensitive to spectral peaks than spectral valleys (Schroeder 1982). This has also been demonstrated in psychoacoustic listening tests, see, e.g., (Moore, Oldfield & Dooley 1989) for a review. Hence, an allpole model is useful not only because it may be a *physical* model for a signal, but because it is a *perceptually* meaningful parametric representation for a signal. In *frequency-warped* LPC, WLPC, an allpole model has a modified frequency representation approximating the frequency representation of human hearing. The main proposition of this thesis is that a warped linear predictive model leads to a perceptually meaningful and efficient parametric representation of audio and speech signals.

Modern audio and speech coding algorithms are based on utilization of frequency masking properties of human hearing (Schroeder, Atal & Hall 1979). Computational models for frequency masking are based on a spectral representation of a signal, for example, in (Karjalainen 1985, Beerends & Stemerdink 1996, Brandenburg & Sporer 1992). Therefore, the design of a perceptual subband codec is relatively straightforward in the sense that perceptual modeling can be incorporated to the algorithm in a natural and intuitive way.

[12] This approach is also called Transform Predictive Coding, TPC (Chen & Wang 1996, Ramprashad 1999).

[13] More complex *source models* for music signals have been recently studied, e.g., in (Tolonen 2000)

1.3 Contents of this thesis

This doctoral thesis consists of a summary and nine articles. The articles are related to frequency-warped signal processing and warped linear predictive coding techniques for audio signals. In Chapter 2, an introduction to classical linear predictive techniques is given. Chapter 3 focuses to the contribution of this thesis. In Chapter 4, the contribution of the current author in the development of the presented techniques is clarified and the main results of each article are summarized. This is followed by an errata for the publications, and copies of the included articles.

2. Theoretical background

2.1 Linear stationary signal models

A discrete signal is a sequence of samples

$$x(n), \text{ where } n = \cdots, -2, -1, 0, 1, 2, \cdots \tag{2.1}$$

Signal $x(n)$ can always be expressed as a linear combination of a set of some other sequences

$$x(n) = \sum_{k=1}^{L} c_k s_k(n). \tag{2.2}$$

For example, in

$$x(n) = \sum_{k=1}^{L} c_k e^{i2\pi kn/L}, \tag{2.3}$$

the signal is expressed as a linear combination of complex exponentials. If $n = 1, 2, \cdots, L$, this is called the *inverse discrete Fourier transform*. The basis functions of this decomposition are defined by the following formula:

$$s_k(n) = e^{i2\pi kn/L}. \tag{2.4}$$

These are a set of orthogonal functions and they form a *complete* basis. This means that a set of L basis functions can represent any signal of duration L exactly. In audio coding applications, subband coding algorithms are based on this principle, that is, critical downsampling with perfect reconstruction.

In theory, signals which can be represented exactly by, e.g., a set of elementary functions, are called *singular* (Wold 1954) or *deterministic* (Doob 1944) signals. A formal definition for a singular signal is that it has a non-continuous power spectrum, see, e.g., (Kailath 1974, Papoulis 1985). Naturally, this definition is not well suited to discrete signals of finite length. Therefore, we call a signal deterministic when it is associated with a deterministic signal model given by 2.2.

The autocorrelation function of a discrete ergodic signal $s(n)$ is defined by

$$R_k = E[s(n)s(n-k)] = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} s(n)s(n-k), \text{ for all } k. \tag{2.5}$$

White noise is a discrete stationary random signal $r(n)$ defined as a sequence with

$$R_k = E[r(n)r(n-k)] = 0, \text{ for all } n \neq k. \tag{2.6}$$

In classical literature (Kolmogorov 1941), $r(n)$ is sometimes called a *fundamental* sequence. In practical applications, signals are of finite length, and therefore a signal may be called random only in respect to some signal model.

2.1.1 Wold decomposition theorem

Signal $x(n)$ can always be written as a sum of a deterministic signal $s_d(n)$ and another signal $s_r(n) = x(n) - s_d(n)$. If $x(n)$ is a stationary signal and $s_d(n)$ and $s_r(n)$ are uncorrelated, it can be shown (Wold 1954) that

$$x(n) = s_d(n) + s_r(n) = s_d(n) + \sum_{k=0}^{\infty} c_k r(n-k), \tag{2.7}$$

where $r(n)$ is an uncorrelated white noise signal and $\sum_{k=0}^{\infty} |c_k|^2 < \infty$. This is called the *Wold decomposition theorem* for a stationary signal[1]. In classical terms (Kolmogorov 1941), $s_r(n)$, which is obtained from a fundamental sequence by *sliding summation*, is called a *regular* sequence. The Wold decomposition is of fundamental importance because it clearly divides the universe of linear spectral estimation methods into two main branches: *deterministic*, and *stochastic* techniques. Deterministic techniques can be associated with *non-parametric* coding techniques such as transform coding. Similarly, *parametric* techniques are usually related to a stochastic signal modeling principle [2].

2.1.2 Prediction problem

If the *coefficients* c_k in (2.7) are fixed and $s_d(n) = 0 \, \forall n$, $x(n)$ is a *moving average* (Slutsky 1927), MA, model for the *stochastic* process $s_r(n)$ given by

$$s_r(n) = \sum_{k=0}^{\infty} c_k r(n-k). \tag{2.8}$$

Clearly, the regular sequence $s_r(n)$ is obtained from white noise $r(n)$ by a convolution with a one-sided infinitely long coefficient sequence c_k, i.e., filtering with an IIR, *Infinite Impulse Response* filter. The Z transform of (2.8) is given by

$$S_r(z) = C(z)R(z) = \left[\sum_{k=0}^{\infty} c_k z^{-k}\right] R(z). \tag{2.9}$$

From (2.9), it is easy to see that

$$R(z) = \frac{S_r(z)}{C(z)} \tag{2.10}$$

[1] Wold decomposition theorem was introduced in the first edition of Wold's book, his doctoral thesis, in 1938. The proof of the theorem can be found in different forms, e.g., in (Wold 1954, Kolmogorov 1941, Priestley 1981, Papoulis 1985).

[2] Nonlinear parametric techniques such as sinusoidal modeling do not fit nicely into this division.

which shows that the white noise *excitation* $r(n)$ is uniquely determined by the filter, its output $s_r(n)$, and the initial conditions at the filters' states.

One may write (2.8) into the following form

$$c_0 r(n) = s_r(n) - \sum_{k=1}^{\infty} c_k r(n-k) = s_r(n) - \tilde{s}_r(n). \tag{2.11}$$

In the following, we simplify notation by denoting $r(n) = c_0 r(n)$, that is, we assume that $c_0 = 1$.

This expression (2.11) has two important aspects. Firstly, $r(n)$ obeys (2.6). Therefore, it also holds that $r(n)$ is uncorrelated with any linear combination of its past values $r(n-k),\ k \geq 1$. That is,

$$E[r(n)\sum_{k=1}^{\infty} c_k r(n-k)] = E[r(n)\tilde{s}_r(n)] = 0. \tag{2.12}$$

This is called the *orthogonality principle*. Secondly, as it was pointed out by Kolmogorov (1941), $\tilde{s}_r(n)$, which is uniquely determined by the history of $s_r(n)$, can be seen as a *linear prediction* for $s_r(n)$. The *prediction error* is, by definition, a white noise signal $r(n)$. Therefore, (2.12) is an optimal solution to the *prediction problem* given by (2.11).

Independently, and in parallel with Kolmogorov's work, Wiener (1949) studied the prediction problem for *continuous signals* from a slightly different perspective [3]. Levinson (1947) extended Wiener's theory for discrete-time signals. They started with minimization of the expectation of (2.11) by

$$\frac{\partial E[|r(n)|^2]}{\partial c_k} = 0,\ \text{where } k = 1, 2, \cdots, \infty \tag{2.13}$$

which leads to the same orthogonality condition given by (2.12) for an *optimal* set of coefficients c_k. It can be shown that this always gives the minimum of the expression, see, e.g., (Levinson 1947). Basically, this is the classical *least squares* regression technique which was already used by Gauss and first published by Legendre in early 19th century, see, e.g., (Kailath 1974, Sorenson 1980, Robinson 1982) for a historical survey. For time series, this technique was first applied by Yule (1927) and Walker (1931).

2.2 Linear prediction

To bring this scheme closer to practical digital signal processing techniques it is next assumed that the Z-transform of an infinite impulse response filter $C(z)$ can be approximated by a finite order rational polynomial, i.e., a finite order IIR filter given by

$$A(z) = \frac{\sum_{m=0}^{K} b_m z^{-m}}{\sum_{p=0}^{L} a_k z^{-k}} \tag{2.14}$$

[3] Wiener (1949) recognizes Kolmogorov's work with the same problem in his book and points out that: ... *the parallelism between them may be attributed to the simple fact that the theory of the stochastic processes had advanced to the point where the study of the prediction problem was the next thing on the agenda.*

In the time domain, (2.8) is now given by

$$s_r(n) = \sum_{m=0}^{K} b_m r(n-m) - \sum_{k=1}^{L} a_k s_r(n-k). \tag{2.15}$$

The first term on the right hand side of (2.15) is a finite order moving average MA, process (Slutsky 1927). In DSP terms, this is an output of a finite impulse response, FIR, filter. The second term, where the value is composed as a weighted combination of past values of $s_r(n)$ is called an autoregressive, AR, process[4] (Yule 1927, Walker 1931), which can be seen as an output of an infinite impulse response, IIR, filter.

In this thesis, the focus is in autoregressive modeling or *linear prediction*[5] and related filtering techniques. Readjusting the notation, a signal $x(n)$, where $n = 0, 1, 2, \cdots N-1$ can be modeled by

$$x(n) = \sum_{k=1}^{L} a_k x(n-k) + e(n) = \tilde{x}(n) + e(n), \tag{2.16}$$

where a_k are the coefficients of an Lth order IIR filter. The *prediction error* signal, or *residual* $e(n)$ may be associated with the random signal $r(n)$ of equations (2.7), (2.12), and (2.13)[6]. In the information theory $e(n)$ is often called the *innovation* sequence (Kailath 1974).

Signal model of Eq. (2.16) is different from that given by (2.2) because the signal is not modeled as a deterministic sequence but as a regular one. In terms of Wold's decomposition, a regular sequence $x(n)$ is obtained from a fundamental sequence $e(n)$ by sliding summation with an impulse response of an IIR filter characterized by the coefficients a_k. As shown above, a signal $x(n)$ and the coefficients a_k can be related to each other by the orthogonality principle of (2.12). In terms of the Wiener's theory, for a signal $x(n)$, the set of optimal coefficients a_k in MMSE sense obeys

$$E[e(n) \sum_{k=1}^{L} a_k e(n-k)] = E[e(n)\tilde{x}(n)] = 0, \tag{2.17}$$

which is a finite-order version of Eq. (2.12).

Using the signal autocorrelation (2.5) and (2.16), one may write (2.17) to the following form, see, e.g., (Levinson 1947, Makhoul 1975, Markel & Gray 1976):

$$\sum_{k=1}^{L} a_k R(p-k) = R(p), \text{ where } p = 1, 2, \cdots, L, \tag{2.18}$$

which is usually called the *Yule-Walker* equations [7].

[4] A famous classical example (Yule & Kendall 1958) of an AR process is the swinging of a pendelum which is pelted by small boys at random with peas.

[5] This name was first used by Wiener (1949). Wold (1954) called this technique as linear autoregression with application to *forecasting*.

[6] In terms of Wold decomposition theorem, if $1/A(z)$ is only an approximation of $C(z)$, the residual $e(n)$ is also a *regular* sequence produced from white noise by filtering with $A(z)C(z)$.

[7] So called *Wiener-Hopf* equations, which are used in solving coefficients for a *Wiener filter*, see, e.g., (Haykin 1996), reduce to *Yule-Walker* equations if the *desired* input signal is the same as the input signal. Some authors call this also Wiener-Hopf equations. This can be motivated if $L \to \infty$. Due to the relation to the orthogonality condition (2.12) these are also called *normal equations*.

This set of linear equations is convenient to express in a matrix form given by

$$\begin{bmatrix} R_0 & R_1 & R_2 & \cdots & R_{L-1} \\ R_1 & R_0 & R_1 & \cdots & R_{L-2} \\ R_2 & R_1 & R_0 & \cdots & R_{L-3} \\ \vdots & \vdots & \vdots & & \vdots \\ R_{L-1} & R_{L-2} & R_{L-3} & \cdots & R_0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_L \end{bmatrix} = \begin{bmatrix} R_1 \\ R_2 \\ R_3 \\ \vdots \\ R_L \end{bmatrix}. \tag{2.19}$$

The matrix on the left hand side of (2.19) is a *Toeplitz* matrix. There are several techniques to solve the coefficients a_k from this matrix expression, see (Makhoul 1975) for review. Levinson (1947) worked out a computationally efficient technique to solve the coefficients a_k. Durbin (1960) introduced a more compact version of this algorithm which is today known as the *Levinson-Durbin* algorithm. This is an *order-recursive* algorithm which utilizes the symmetry of the Toeplitz matrix. The results of computation at previous stages are utilised in following stages. In the standard version of the algorithm, the intermediate results are the same as the *reflection coefficients* of a corresponding lattice filter, see Section 2.2.3. Even more efficient variations of this algorithm have been introduced such as the *split*-Levinson algorithm (Delsarte & Genin 1986).

In (2.18), the autocorrelation function for a signal of infinite duration given by (2.5) was adopted even if the length of the signal in any practical case is finite. This mismatch between theoretical concepts and practical digital signal processing methods exists in the light of the *Wiener-Kolmogorov* theory (Åström 1970, Priestley 1981). Most of the problems could be avoided using more elaborate theory. In particular, so called *Kalman-Bucy* theory (Kalman 1960) extends the theory of optimal prediction and filtering for signals of finite length, see, e.g., (Kailath 1974, Haykin 1996), for review. However, this extension is omitted in this thesis.

The autocorrelation function in (2.18) can be interpreted as that of an infinitely long signal which is windowed so that it is non-zero only in the range of interest. Several different window functions can be used with this including classical choices such as a rectangular window or the Hamming window (Blackman & Tukey 1959). This approach is usually called the *autocorrelation method* of linear prediction (Makhoul 1975). In the case of a rectangular window function, correlation terms are computed with

$$R_k = \frac{1}{N}\sum_{n=0}^{N-1} x(n)x(n-k), \text{ where } x(n) = 0 \text{ for all } n < 0 \text{ and } n > N-1. \tag{2.20}$$

Another approach is to change the expectation operator in (2.13) to a finite sum. This gives

$$\frac{\partial \frac{1}{N}\sum_{n=0}^{N-1} |e(n)|^2}{\partial a_k} = 0, \text{ where } k = 1, 2, \cdots, L \tag{2.21}$$

and leads to the following matrix form

$$\begin{bmatrix} C_{0,0} & C_{0,1} & C_{0,2} & \cdots & C_{0,L-1} \\ C_{1,0} & C_{1,1} & C_{1,2} & \cdots & C_{1,L-1} \\ C_{2,0} & C_{2,1} & C_{2,2} & \cdots & C_{2,L-1} \\ \vdots & \vdots & \vdots & & \vdots \\ C_{L-1,0} & C_{L-1,1} & C_{L-1,2} & \cdots & C_{L-1,L-1} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_L \end{bmatrix} = \begin{bmatrix} C_{0,1} \\ C_{0,2} \\ C_{0,3} \\ \vdots \\ C_{0,L} \end{bmatrix}, \tag{2.22}$$

where the correlation terms are given by

$$C_{k,p} = \frac{1}{N} \sum_{n=0}^{N-1} x(n-k)x(n-p). \tag{2.23}$$

This way of formulating the equations in solving coefficients for a linear predictor is called the *covariance method* of linear prediction (Makhoul 1975).

2.2.1 Linear predictive coding

Linear predictive coding, LPC, is an application of linear prediction modeling to signal encoding [8]. For speech coding applications this was proposed in (Atal & Schroeder 1967, Atal & Hanauer 1971, Itakura & Saito 1970). *Prediction error* form of LPC encoder [9] follows directly writing (2.16) to the following form:

$$e(n) = x(n) - \sum_{k=1}^{L} a_k x(n-k) \tag{2.24}$$

The Z transform of (2.24) is given by

$$E(z) = X(z)\left(1 - \sum_{k=1}^{L} a_k z^{-k}\right) = X(z)A(z), \tag{2.25}$$

where $A(z)$ is called the *prediction error filter*, or *inverse filter*. In using the autocorrelation method of linear prediction this is a minimum-phase finite impulse response, FIR, filter, see, e.g., (Haykin 1989) for the derivation of this property.

The encoding process involves computation of filter coefficients a_k and the prediction error signal, or the residual $e(n)$. In the decoder, the original signal is reproduced using

$$X(z) = \frac{E(z)}{1 - \sum_{k=1}^{L} a_k z^{-k}} = \frac{E(z)}{A(z)}, \tag{2.26}$$

where $1/A(z)$ is now called the *synthesis filter*, which is a minimum-phase infinite impulse response, IIR, filter.

The residual $e(n)$ and filter coefficients a_k must be transmitted to the decoder, that is, they should be quantized[10]. Several papers have been published about different strategies

[8] *Predictive coding* is usually associated with early articles by Cutler (1952) and Elias (1955).

[9] Prediction error coder, PEC (Gibson 1980), is also called *predictive-subtractive coder* (Oliver 1952) and D*PCM (Noll 1975).

[10] See, e.g., (Gray & Neuhoff 1998) for an extensive literature survey on quantization

for quantization of LPC coefficients, see (Viswanathan & Makhoul 1975, Paliwal & Kleijn 1995), for details. The quantization of the residual, or excitation for the synthesis filter, can be based either on scalar quantization (Jayant 1973, Noll 1975, Noll 1978), see (Jayant & Noll 1984), for review, or vector quantization, e.g., in (Schroeder & Atal 1985, Gersho 1994, Kroon & Kleijn 1995) [11]. In the following, quantization is denoted by a quantizer operator $Q[\cdot]$.

The quantization error $e_q(n) = e(n) - Q[e(n)]$, or in the Z-domain $E_q(z)$, is typically a nearly white noise process (Jayant & Noll 1984). Since $1/A(z)$ is a linear filter, its output for $E(z) + E_q(z)$ can be expressed as a sum of a clean signal $X(z)$ and an additive noise signal $X_q(z)$. In terms of Eq. (2.26) we have

$$X(z) + X_q(z) = \frac{E(z)}{1 - \sum_{k=1}^{L} a_k z^{-k}} + \frac{E_q(z)}{1 - \sum_{k=1}^{L} a_k z^{-k}}, \tag{2.27}$$

Equation (2.27) states that the *coding error* signal in prediction error coding has the Z-transform, $X_q(z)$, characterized by the estimated allpole model $1/A(z)$. Reformulating (2.25) to

$$E(z) = X(z)\left(1 - \sum_{k=1}^{L} a_k z^{-k}\right) - E_q(z)P(z), \tag{2.28}$$

we have a *closed-loop* encoder. Typically, $P(z) = 1 - \sum_{k=1}^{L} \gamma^k a_k z^{-k}$, where *bandwidth expansion parameter* $0 < \gamma < 1$. If $\gamma = 1$, $X_q(z) = E_q(z)$, that is, coding error is approximately white noise [12]. Other choices for $P(z)$ can be used to shape the spectrum of the coding error signal (Makhoul & Berouti 1979).

Levinson's work (Levinson 1947) with discrete linear predictive *deconvolution*, that is, inverse filtering, was first applied to an engineering application, analysis of seismic data for oil industry, by Tukey in 1951, see (Robinson 1982). The technique soon established its position especially in exploration of oil and natural gas. This success boosted the development of the theory and the techniques of LPC at that field. In particular, Burg's *maximum entropy* spectral analysis technique (Burg 1967, Burg 1975) has gained attention also in speech coding applications (Schroeder 1982). Here, instead of minimizing the energy of the prediction error, the goal is to maximize the entropy. In terms of classical measures for the performance of a LPC model, see, e.g., (Jayant & Noll 1984), the former is based on maximization of *prediction gain*, while the latter tries to maximize the *spectral flatness*. Makhoul (1977) showed that Burg's method is equivalent with a certain formulation of a *lattice method* of linear prediction, which is discussed below.

2.2.2 Spectral representation

Itakura & Saito (1970) developed [13] a mathematical model for speech power spectrum based on a *maximum-likelihood* matching of speech power spectrum $P(\omega)$ with a para-

[11] In this context LPC is usually called *Code Excited Linear Prediction*, CELP.

[12] As was the case in (Atal & Schroeder 1967)

[13] Markel & Gray (1976) cited to a Japanese internal report by Saito and Itakura published in 1966.

metric allpole spectrum model given by

$$\tilde{P}(z) = \frac{\sigma_e^2}{2\pi} \frac{1}{|1 - \sum_{k=1}^{L} a_k z^{-k}|^2}, \tag{2.29}$$

where σ_e^2 is a scale factor for the magnitude. Assuming *Gaussian* distribution for the input signal, their approach also leads to the autocorrelation method of linear prediction introduced above. The autocorrelation function and the power spectrum of a stationary signal form a Fourier transform pair (Wiener 1930, Khintchine 1934, Wold 1954)[14]. The spectral theory of autoregressive modeling, or linear prediction, was already established in (Whittle 1954). In the spectral domain the minimization of the square of the prediction error in (2.13) is equivalent to minimizing

$$E_{\mathrm{LP}} = \frac{\sigma_e^2}{2\pi} \int_{-\pi}^{\pi} \frac{P(\omega)}{\tilde{P}(\omega)} d\omega. \tag{2.30}$$

Taking the logarithm of the integrand[15] we have

$$E_{\mathrm{LPlog}} = \frac{\sigma_e^2}{2\pi} \int_{-\pi}^{\pi} \log\left(\frac{P(\omega)}{\tilde{P}(\omega)}\right) d\omega = \frac{\sigma_e^2}{2\pi} \int_{-\pi}^{\pi} \log(P(\omega)) - \log\left(\frac{1}{|1 - \sum_{k=1}^{L} a_k e^{-i\omega}|^2}\right) d\omega. \tag{2.31}$$

The role of LP on a log-magnitude power spectral domain is to minimize the mean square difference between the logarithmic power spectrum of a signal and a corresponding log-magnitude allpole spectrum[16].

The inverse Fourier transform of the power spectrum is the autocorrelation function which can be used to compute the coefficients of an allpole filter using (2.19). Using this path in estimating coefficients a_k, it is also possible to incorporate various types of frequency domain criteria into the process. A classical example is *selective linear prediction* (Makhoul & Cosell 1976). Here, two regions of spectrum are considered separately and a model of different order is applied to them (Makhoul 1975, Markel & Gray 1976). Another example is the *Perceptual Linear Prediction* by Hermansky (1990) where the all-pole model is fitted to a *loudness* spectrum on the psychoacoustic Bark scale (Scharf 1970). In these two examples the linear predictive machinery is typically used only for signal analysis, because the implementation of filters in (2.25) and (2.26), for analysis and synthesis, is difficult or impossible. In addition, these techniques are based on computation of the power spectrum, which is typically done using non-parametric spectral estimation techniques such as the Fast Fourier Transform, FFT, (Cooley & Tukey 1965). This indirect way of getting the *correlation coefficients* for (2.19) may make the technique computationally expensive and sensitive to errors. This thesis studies certain versions of selective linear prediction where there is a direct implementation for the prediction error and synthesis filters, and the correlation terms can be computed directly from the input signal.

[14] This is sometimes called the *Einstein-Wiener-Khintchine* theorem.

[15] Power spectra of regular sequences are positive and non-zero everywhere, see, e.g., (Papoulis 1985).

[16] Imai & Furuichi (1988) have introduced an interesting technique where an unbiased *cepstral* coefficients are estimated from log-magnitude, or log-generalized (Kobayashi & Imai 1984), spectrum. The obtained generalized cepstral coefficients can be used directly with log-magnitude approximation filters (Imai 1980).

2.2.3 Lattice methods

It is possible to convert any digital filter to a corresponding lattice filter (Itakura & Saito 1972). The coefficients of lattice filters are called *reflection coefficients*[17] (Makhoul & Cosell 1976).

Reflection coefficients have many interesting properties. In (Atal & Hanauer 1971), these coefficients were derived directly from a non-uniform acoustic tube model, where the coefficients, as the name indicates, are reflection coefficients of individual tube elements. Therefore, the reflection coefficients and the lattice structure have firm physical interpretations. Their goal was to find a representation for LPC coefficients which is more robust to quantization [18]. Reflection coefficients also act in a reasonable way in temporal interpolation of coefficients between frames. In addition, if all the reflection coefficients obey $|K_p| < 1,\, p = 1, 2, \cdots, L$, the synthesis filter is stable. Therefore, lattice methods of linear prediction also give direct means to check and guarantee the stability of the estimated model.

Itakura & Saito (1972) introduced a technique to estimate the reflection coefficients directly from *forward* and *backward* prediction error signals in the lattice structure. Makhoul (1977) proposed a class of lattice methods for linear predictive modeling which comprises also Itakura's and Burg's methods (Burg 1975) as special cases. Friedlander (1982) further extended this work by introducing a large set of alternative techniques for time-invariant and also time-varying spectral modeling using lattice filter structures. Lattice methods of linear prediction are *order-recursive*. That is, the optimal coefficients are first solved for the first stage of the filter, then the prediction error signals are computed for the next stage and so on.

2.3 Linear nonstationary signal models

The techniques and concepts discussed above are all based on an assumption about *stationarity* of the input signal. In practical LPC algorithms, the filter coefficients are time-varying, i.e, parameters of a *nonstationary* signal model. The basic technique to obtain this is to perform linear predictive analysis in frames such that the signal is assumed to be stationary within each analysis frame. In a long time scale, this means that the signal model for linear predictive coding, from (2.16), is actually given by

$$x(n) = \sum_{k=1}^{L} a_k(n)x(n-k) + e(n) = \tilde{x}(n) + e(n), \tag{2.32}$$

where filter coefficients $a_k(n)$ are now also functions of time n. This is called a nonstationary signal model for linear prediction.

Booton (1952) extended Wiener's (Wiener 1949) theory to nonstationary signals and Cremer (1961) showed that Wold decomposition principle applies also to nonstationary

[17] On the field of statistics, these are called *partial correlation* coefficients (Box & Jenkins 1970, Priestley 1981).

[18] However, there are more favorable representations available today (Paliwal & Kleijn 1995).

signal models. It is possible to formulate an orthogonality condition of Eq. (2.12) for this signal model. However, there is no unique solution for the optimal time-varying coefficients $a_k(n)$. The *coefficient evolutions* must be restricted somehow in order to find one of the least-square optimal solutions to the coefficients.

It is usually assumed that the signal is *locally stationary* (Makhoul 1975), or the coefficients are *smoothly time-varying*, see, e.g., (Priestley 1981). These are conceptually two different approaches. The local stationarity assumption is used in conventional frame-based and continuously adaptive techniques. Smooth coefficient evolution is assumed in *smoothness priors* techniques (Kaipio & Juntunen 1999, Juntunen 1999), and in techniques where the coefficient evolutions are restricted to a class of functions which can be expressed as linear combinations of predefined basis functions(Subba Rao 1970). The latter approach is sometimes called *deterministic regression* approach of time-varying autoregressive modeling.

2.3.1 Frame-based processing

Audio and speech coding algorithms usually process the input signal in frames. For example, in LP-based speech coders, the frame-length is typically 10-20 milliseconds. The frames are usually overlapping and, in the case of the autocorrelation method, some window function is applied to each signal frame before analysis. The filter coefficients corresponding to each frame are coded and transmitted along with excitation data. A direct application of this procedure would produce discontinuities to the coefficient trajectories in frame borders, which may produce unwanted artifacts. It is a common practice to *interpolate* filter coefficients smoothly from one frame to another. Therefore, the signal model which is considered in these algorithms is essentially given by Eq. (2.32) even if the spectral model is estimated in locally stationary frames.

The direct-form coefficients of a synthesis filter are not a convenient representation for coefficient interpolation. Therefore, interpolation is usually done for reflection coefficients, Log-Area-Ratios, LARs (Viswanathan & Makhoul 1975), or Line Spectrum Frequency, LSF (Itakura 1975), terms [19]. However, this is a somewhat arbitrary approach and not related to the actual fluctuation in the input signal.

It is possible to increase the amount of overlapping in the analysis so that the coefficients are estimated more frequently. An extreme example is a *sliding window* formulation of linear predictive modeling where coefficients are solved at each time instant. However, this is computationally expensive and leads to an increased number of filter coefficients to be transmitted. Barnwell (1977) has introduced a computationally efficient method for computation of *adaptive autocorrelation*. In this method, the correlation terms in (2.19) are computed recursively using a leaky integrator. This is a version of the autocorrelation method of linear prediction where the window function is actually defined as an impulse response of a low-order IIR filter.

[19] This is usually done already for efficient quantization of coefficients.

2.3.2 Adaptive filtering

Stochastic gradient methods[20] for adaptive filtering also follow from a *local* formulation of the prediction problem. Here, the coefficients are not solved directly for a long signal frame but adjusted iteratively such that the filter coefficients converge, in the case of a stationary signal, towards optimal values. In this sense, these techniques are *time-recursive*. A classical example is Least-mean-square, LMS, algorithm which was presented by Widrow & Hoff Jr. (1960). In the LMS algorithm, and its many variants (Haykin 1996), the coefficients of a direct form filter are adjusted using a simple gradient rule. Gradient adaptive lattice method, GAL, by Griffiths (1977) is an application of the same principle to a lattice filter. Due to the cascaded structure of a lattice filter, the GAL algorithm is both time-recursive and order-recursive. In practice, GAL algorithm is significantly faster in convergence than the conventional LMS algorithm (Haykin 1996).

In adaptive filtering techniques, the gradient update rule can also be interpreted as a method to produce a recursive window function for linear predictive analysis.

Adaptive filtering techniques are not directly suitable for coding applications because they produce a set of filter coefficients at each sample which should be coded and transmitted to the receiver. Gibson, Jones & Melsa (1974) introduced a *backward adaptive* formulation of linear predictive coding. This is a close relative to backward adaptive quantization methods presented, e.g., in (Jayant 1973). Here, the spectral model is not formed from the original input signal but from the already coded and transmitted signal. Since the same model can be computed at the decoder, there is no need to code and transmit filter coefficients. However, the spectral model is completely estimated from the signal already transmitted. Therefore, the coefficients should be updated very frequently. Several different adaptive filtering techniques were compared in (Gibson, Cheong, Chang & Woo 1990).

Backward adaptive linear predictive techniques are especially suitable for *low-delay* coding of speech and audio signals. Several formulations of this scheme have been proposed, see, e.g., (Chen 1995), for review. Iyengar & Kabal (1988) introduced a low-delay speech codec which is based on a backward adaptive formulation of the GAL algorithm, see also (Yatrou & Mermelstein 1988). A *low-delay CELP* algorithm for low-delay speech coding proposed by Chen, Cox, Lin & Jayant (1992) was standardized by ITU-T as the Recommendation G.728[21]. This algorithm is based on a backward adaptive formulation linear prediction where the spectral model is estimated using a modified version of Barnwell's (Barnwell 1977) adaptive autocorrelation method, so called *hybrid windowing* technique (Chen, Lin & Cox 1991).

[20] Or, *steepest decent* methods.

[21] See (Chen & Cox 1993) for an interesting inside story of the work.

2.3.3 Deterministic regression time-varying LPC

It was proposed by Subba Rao (1970) that the time-varying coefficient evolutions $a_k(n)$ could be expressed by

$$a_k(n) = \sum_{\ell=0}^{M} c_{k\ell}\phi_\ell(n), \tag{2.33}$$

where $\phi_\ell(n)$ are a set of M predefined *basis functions*. For this system it is possible to formulate normal equations where the least squares optimal coefficients $c_{k\ell}$ can be solved directly. For speech applications this has been studied, e.g., by Liporace (1975), Hall, Oppenheim & Willsky (1983). Grenier (1983) introduced a similar technique based on a lattice formulation. Typically, basis functions are some elementary mathematical functions such as the Fourier basis, Gaussian pulses, or prolate spheroidal sequences (Slepian 1978).

3. Contributions of this thesis

3.1 Generalized predictor structures

In the conventional form, linear prediction of a current sample is given as a linear combination of each previous value. The Z-transform of the predictor appears in (2.25). This scheme may be generalized by replacing the unit delay elements z^{-k}, for $k = 1, 2, \cdots, L$ with another set of filters $D_k(z)$ [P7]. This gives the following prediction error filter:

$$E(z) = X(z) \left(1 - \sum_{k=1}^{L} a_k D_k(z)\right) = X(z)A(z). \tag{3.1}$$

Correspondingly, the synthesis filter is given by

$$X(z) = \frac{E(z)}{1 - \sum_{k=1}^{L} a_k D_k(z)} = \frac{E(z)}{A(z)}. \tag{3.2}$$

In terms of the Wold decomposition (2.7) this changes nothing, that is, a *regular* sequence $x(n)$ is obtained from residual $e(n)$ using a linear filter $1/A(z)$. However, in terms of the orthogonality principle, it is reasonable to require that the outputs of $D_k(z)$ should be linearly independent. The *gain term* c_0 in (2.11) is usually not unity as in the case of conventional linear prediction. In [P7] this gain term is denoted by g.

Linear predictive coding with modified predictor structures [1] was studied in [P7] in detail. It was shown how correlation terms can be computed and the corresponding normal equations can be formulated and solved. The article shows this general form, but focuses on the case where the *subfilters* $D_k(z)$ form, or approximate, a cascade structure, that is

$$D_k(z) = D_1^k(z), \tag{3.3}$$

where $D_1(z)$ is some *prototype* block whose Fourier transform is given, or approximated by

$$D_1(\omega) = A(\omega)e^{-i\psi(\omega)}. \tag{3.4}$$

For this predictor structure, it can be shown [P7] that the spectrum representation for prediction error in (2.30) takes the following form:

$$E_{\mathrm{LP}} = \int_{-\pi}^{\pi} |X(\omega)|^2 \left| 1 - \sum_{k=1}^{L} a_k A^k(\omega) e^{-ik\psi(\omega)} \right|^2 \left(\frac{\partial \psi(\omega)}{\partial \omega} \right) d\omega. \tag{3.5}$$

[1] which was partly inspired by an article by Laine (1995).

This equation shows that the power spectrum is modeled on a *warped* frequency scale determined by the function $\psi(\omega)$. The magnitude term $A(\omega)$ can be used for spectrum shaping. However, mainly due to stability problems associated with the generalized synthesis filter (3.2), only the cases where the subfilters are allpass filters, i.e., $A(\omega) = 1$, are exemplified in the article. The last differential term in (3.5) produces a spectral tilt to the spectrum.

The first example of the article is a linear predictive codec where $D_k(z)$ are *fractional delay* filters (Laakso, Välimäki, Karjalainen & Laine 1996). This type of an LPC algorithm can be designed so that the LP modeling focuses to a low frequency band of the input signal and completely neglects the spectral information above a certain frequency limit. This is related to the works of Makhoul (1975) with selective linear prediction, where the same effect was achieved using a frequency domain approach. The main advantage of the technique in [P7] is that the filter coefficients can be estimated directly from the waveform and the corresponding prediction error and synthesis filters can be implemented. Moreover, with a suitable selection of $D_k(z)$, one can implement a linear predictive codec where the frequency resolution approximates closely a logarithmic frequency scale.

The article [P7] also reviews earlier work where $D_k(z)$ are a set of orthogonal polynomial functions, see, e.g., (Ninness & Gustafsson 1997) given by

$$D_k(z) = \frac{\sqrt{1 - |\lambda_k|^2}}{1 - \lambda_k z^{-1}} \prod_{p=1}^{k} \frac{z^{-1} - \lambda_p}{1 - \lambda_p z^{-1}}. \tag{3.6}$$

If $\lambda_k = \lambda_p = 0, \forall k, p$ this reduces to a conventional LP predictor. If $\lambda_k = \lambda_p, \forall k, p$ this is so called *Laguerre model* (Lee 1960, King & Paraskevopoulos 1977, Oliveira e Silva 1995*a*). With a suitable selection of parameters in Equation (3.6), this also leads to Kautz models (Kautz 1954, Wahlberg 1994). An extensive literature review on the use of orthogonal subfilters were recently given in (Paatero 2000).

This type of modifications to the prediction scheme have a long tradition. The use of Laguerre functions was already proposed by Lee (1933) and Wiener (1949) in the case of continuous-valued systems. Their work with various types of orthogonal functions was reviewed and extended in (Lee 1960). King & Paraskevopoulos (1977) introduced a discrete version of *Laguerre filter*[2] based on discretized Laguerre functions (Gottlieb 1938). Autoregressive modeling based on discrete Laguerre functions has been studied, e.g., in (Wahlberg 1991, Oliveira e Silva 1995*b*), especially in the field of *system identification*[3] in control theory.

Recently, Varho & Alku (1999) and Chang, Cheong, Ting & Tam (2000) have proposed modified linear predictive structures where a prediction is formed by grouping or selecting past signal samples in different ways. These techniques are obviously related, but they are largely omitted in this thesis. See, e.g., (Varho 2001) for a review.

[2] Their application examples were a low-order filter with a triangular impulse response and a Hilbert transformer.

[3] That is, parametric modeling.

Most of the articles in this thesis deal with systems where

$$D_k = \left(\frac{z^{-1} - \lambda}{1 - \lambda z^{-1}}\right)^k . \tag{3.7}$$

In this case, the filters $A(z)$ and $1/A(z)$ are called *warped* FIR and IIR filters, respectively[4] The difference between a warped filter and a Laguerre filter is that the latter has an additional pre-filter, see (3.6) given by

$$W(z) = \frac{\sqrt{1 - \lambda^2}}{1 - \lambda z^{-1}}. \tag{3.8}$$

The role of $W(z)$ is to orthogonalize the set of filters. In many practical applications this is just an additional lowpass filter for a warped filter and therefore the difference between a warped filter and a Laguerre filter is insignificant.

The implementation of the generalized synthesis filter given by (3.2) is not necessarily a straightforward task. Let us study a simple first-order system with $D_1(z) = 1 - z^{-1}$. This yields

$$X(z) = \frac{E(z)}{1 - a_1(1 - z^{-1})}, \tag{3.9}$$

which has the following difference equation

$$x(n) = e(n) + a_1 x(n) + a_1 x(n-1), \tag{3.10}$$

The output $x(n)$ of the filter appears on the both sides of the equation. That is, the filter has a *delay-free* loop structure which can not be implemented directly[5]. The solution in this case is trivial:

$$x(n) = \frac{e(n)}{1 - a_1} + \frac{a_1 x(n-1)}{1 - a_1} \tag{3.11}$$

Equation (3.11) shows an equivalent filter where the delay-free loops have been eliminated. This can be implemented directly if $a_1 \neq 1$[6]. In the case of more complex filter structures it is a significantly more challenging task to make this modification (Szczupak & Mitra 1975, Toy & Chirlian 1984, Karjalainen, Härmä & Laine 1996). In any case, the modified filter is a new filter with another set of filter coefficients. If the original coefficients a_k are obtained, for example, using modified linear prediction, the coefficients of the *realizable* filter must be computed each time the coefficients are changed. For example, in continuously adaptive filtering (Haykin 1996) or continuous interpolation of filter coefficients this mapping from the coefficients of the original filter to those of the realizable structure must be done at each sample. Typically this is a computationally expensive task.

In (Härmä 1998*b*) and [P6], the author developed a new approach for the implementation of recursive generalized filters[7]. Two different techniques are introduced in [P6].

[4] The terminology is inaccurate. Since $D_1(z)$ in (3.7) is an IIR filter, $A(z)$ is also an IIR filter. The name warped FIR, WFIR, is used to illustrate the structural similarity to the conventional FIR filter. In some of the articles, e.g., [P6], these are called warped *all-zero* and *all-pole* filters, respectively. This is also misuse of terminology because both filters are actually pole-zero- filters, see, [P7].

[5] One should know $x(n)$ in order to compute its value.

[6] If $a_1 = 1$ is substituted into (3.9), the constant term vanishes, and the filter becomes *non-causal*

[7] In fact, (Härmä 1998*b*) studies this from a generalized view and gives examples of warped IIR filters, while [P6] focuses to the implementation of warped IIR filters. The derivation of the technique is more accurate in [P6].

Firstly, there is an algorithm which can be used to implement directly any recursive filter having the transfer function given by $1/A(z)$. This is based on splitting the implementation into two steps where the output of the filter is first computed, after which, its inner states are updated. The algorithm makes it possible to implement a filter without changing the structure or the coefficients of the filter. Secondly, the derivation of this technique also gives a generic procedure for modifying a filter structure so that the delay-free loops are eliminated. In the case of a warped IIR filter, this approach leads to exactly the same modified structure as was presented earlier in (Imai 1983, Karjalainen et al. 1996). The availability of these two techniques made it possible to implement various types of warped direct-form and lattice synthesis filters in [P2], [P3], [P4], [P7], and [P8].

3.2 Frequency-warped signal processing

Oppenheim et al. (1971) introduced a technique to compute FFT with a non-uniform spectral resolution using the outputs of a chain of first-order allpass filters, see [P8], for examples. Most of the fundamental properties of frequency-warped signal processing were already introduced in (Oppenheim & Johnson 1972) and (Braccini & Oppenheim 1974). The phase function of a first-order allpass filter, given by $D_1(z)$ in (3.7), is

$$\psi(\omega) = \omega + 2\arctan\left(\frac{\lambda\sin(\omega)}{1-\lambda\cos(\omega)}\right), \tag{3.12}$$

where λ is called here a *warping parameter* [8]. As discussed above, the phase function of a subfilter in (3.5) produces a non-uniform frequency resolution for the LP system. The same frequency-warping effect also occurs in computing the FFT over the outputs of an allpass filter chain. The frequency-warping effect is studied in [P9]. A real-valued λ produces the best frequency resolution at low or at high frequencies, depending on the sign of the parameter, $\lambda > 0$ or $\lambda < 0$, respectively[9]. It is shown how different classes of digital signal processing algorithms can be *warped* by replacing the unit delays of a conventional filter by first order allpass filters and how this yields systems with a warped frequency representation.

Constantinides (1970) introduced techniques for spectral transformations for digital filters by means of replacing the unit delays of a conventional structure by first or second order allpass subfilters. This type of filter transformations for a lattice filter has been studied by Messerschmitt (1980). Based on filter transformations, Schüssler (1970) introduced a *variable digital filter* where the cutoff frequency of a transformed filter could be adjusted by varying a single parameter, that is, the warping parameter λ. This approach has been used by many authors, e.g., in (Johnson 1976, Li 1998).

It was pointed out by Strube (1980) that the frequency mapping in a warped system is relatively close to that of human hearing if the warping parameter λ is chosen appropri-

[8] This is called *discount factor*(King & Paraskevopoulos 1977) or *Laguerre parameter* in the case of a Laguerre filter.

[9] It is also possible to use a complex-valued λ parameter as was proposed in (Oppenheim & Johnson 1972) and [P1] to place the range of best resolution more freely.

ately. Smith & Abel (1999) derived an analytic expression [10] for optimal value of λ such that the frequency resolution of a warped system approximates that of human hearing, e.g., the Bark scale (Scharf 1970, Zwicker & Fastl 1990) or the ERB rate-scale (Moore, Peters & Glasberg 1990). Recently, den Brinker (1998) has given an interpretation for the critical bands of hearing in terms of a local Kautz transformation.

The frequency scales of human hearing are reviewed and compared with frequency-warped frequency representation in [P9]. In addition, [P9] introduces a number of audio applications where the use of warped filters have shown advantages over conventional systems mainly due to better match with the frequency resolution of hearing. Typical applications are design of Bark-scale filterbanks (Laine & Härmä 1996, Evangelista & Cavaliere 1998, Sarroukh & den Brinker 1998), filters for loudspeaker equalization (Karjalainen, Piirilä, Järvinen & Huopaniemi 1999, Asavathiratham, Beckmann & Oppenheim 1999, Pedersen, Rubak & Tyril 1999), HRTF filtering (Huopaniemi, Zacharov & Karjalainen 1999), and modeling of musical instruments (Karjalainen & Smith 1996). The warped FIR and IIR filters can be designed using basically any conventional time-domain or frequency-domain method for filter design. In using time-domain methods the impulse response of the filter must be first warped. For Laguerre FIR filters this technique was introduced by Maione & Turchiano (1985) and for warped filters by several authors cited in [P9]. In using frequency-domain techniques the frequency response of the filter must be specified on a warped frequency scale, see, e.g., (Karjalainen, Härmä & Laine 1997).

3.3 Warped linear prediction

Warped linear prediction, WLP, was first introduced by Strube (1980)[11] The technique was applied to speech coding in (Krüger & Strube 1988). For wideband audio applications this technique was used in (Laine, Karjalainen & Altosaar 1994)[12]. This article also introduced an efficient technique for the computation of warped autocorrelation function, i.e., the warped autocorrelation network. A slightly different approach for WLP was recently introduced in (Edler & Schuller 2000), where a related technique was applied for adaptive pre- and post-filtering in a wideband audio codec.

A group of researchers, e.g., in (Tokuda, Kobayashi & Imai 1995, Koishida, Tokuda, Kobayashi & Imai 1996, Koishida, Hirabayashi, Tokuda & Kobayashi 1998), has systematically employed their *mel-generalized cepstral* techniques (Tokuda, Kobayashi, Imai & Chiba 1993) for speech analysis, coding, and synthesis. WLP technique (Strube 1980) can be seen as a special case of their generalized scheme which also incorporates classical homomorphic (Oppenheim & Schafer 1968) cepstral and mel-cepstral techniques (Imai 1983).

[10] This is slightly different and more correct version of the derivation presented in their earlier paper (Smith & Abel 1995). However, there are some typing errors in (Smith & Abel 1999). The correct version of this formula is given, e.g., in [P8].

[11] Strube (1980) refers to earlier works with selective LP (Makhoul & Cosell 1976) already introduced in previous sections.

[12] Laine's publications on warped signal processing are related to his theory of classes of orthonormal FAM and FAMlet functions (Laine 1992). This is also the theoretical frame of reference in the Master's Thesis (Härmä 1997), Licentiate's Thesis (Härmä 1998*a*), and several earlier publications of the current author (Härmä, Laine & Karjalainen 1997).

The author of this thesis published his first article on warped linear prediction in audio coding in 1996 (Härmä, Laine & Karjalainen 1996). It presented an implementation of a warped prediction error coder with adaptive scalar quantization of the residual signal. This compared the performance of warped LPC and the classical LPC in terms of conventional technical measures such as *prediction gain* and *spectral flatness* (Jayant & Noll 1984). In [P8], this analysis was repeated and it turned out that the difference between the two cases in terms of classical measures is relatively small[13]. A new technical measure which is based on the ability of an estimated model to separate two spectrum peaks is also introduced [P8]. This measure clearly shows the advantages of WLP in respect to frequency resolution of human hearing.

In another conference article, (Härmä et al. 1997), the goal was to study how well a warped LPC could work automatically as a perceptual audio coder. Recall from previous discussion that in a prediction error coder, the spectral shape of a *coding error signal* is close to that of estimated allpole spectral model. When the model is estimated directly on a Bark-warped frequency scale the allpole model can be seen as an approximation of the psychoacoustic *frequency masking pattern* for a complex wideband signal. Therefore, a simple WLPC performs in a somewhat similar way with more complex perceptual audio codecs based on subband decomposition and spectral quantization controlled by a separate auditory model. This was illustrated in (Härmä et al. 1997) by comparing the spectrum of a coding error signal in WLPC and in a MPEG I layer 3 codec[14].

The principle of simplifying the structure of a codec such that the perceptual model is integrated into the coding process was taken even further in [P1]. Here, the two channels of a *stereophonic* audio signal are converted to a single complex-valued signal. The paper presents three alternative techniques for this. The most successful one appeared to be a technique where the signals are converted to analytic signals using the Hilbert transform, the left channel is complex-conjugated and the signals are added. As a result, the signal in the right channel is completely mapped to the right hand side ($[0, \pi]$) of the complex-valued nonsymmetrical spectrum of the obtained complex-valued signal. Correspondingly, the left channel appears on the left hand side of the spectrum that is, at *negative* frequencies ($[-\pi, 0]$).

Linear predictive coding process can be directly derived for complex-valued input signals and filters, see, e.g., (Haykin 1989). This also works with warped LPC, and hence it was possible to formulate a complex-valued warped prediction error coder in [P1] which is driven by a complex-valued stereo signal. There are a number of advantages and disadvantages in this scheme.

A model of fixed order is optimized simultaneously for both channels of the stereo signal. For example, if there is a signal frame where the left channel is almost silent or noisy, most of the poles run to the positive frequencies, e.g., upper half of the unit disc in the Z domain, to model the right channel. This is obviously a favorable way to share the resources in stereo coding. The inverse filter, when working properly, *whitens* the two-

[13] In (Härmä et al. 1996), the authors were unaware of the *gain* and *spectral tilt* factors associated with WLP. Therefore the results indicated a significant difference between the two cases. This was corrected and explained in [P8].

[14] One of the figures from (Härmä et al. 1997) is replotted in [P9] as Figure 19.

sided nonsymmetrical spectrum of the stereo signal. Consequently, the complex-valued residual has almost a symmetrical spectrum and therefore it can be replaced by a single real-valued sequence, e.g., the sum of the real and imaginary parts of the original residual.

The main disadvantage in complex valued processing of a stereo signal is the mapping from two signals to a single complex-valued by means of the Hilbert transform. In (Härmä 1998*a*), the author developed a number of alternative techniques for this mapping. However, no single technique which would be acceptable in terms of processing delay, computational complexity, and accuracy in terms of magnitude and phase[15] was found. This scheme for stereo coding was used in (Härmä, Vaalgamaa & Laine 1998) but it was omitted in further developments of this coding scheme (Vaalgamaa, Härmä & Laine 1999).

The topic of [P8] is comparison of warped LPC to the conventional LPC. This comparison is done using a kind of generalized and simplified prediction error codec. It is assumed in the article that the results obtained with this type of a simulated codec could be extended to cover also more complex modern CELP type speech and audio coding algorithms based on linear prediction. The article reviews the theory of warped linear predictive coding and introduces most of the essential aspects of the technique[16]. The most important part of the article is a report on extensive listening tests which were performed at Helsinki University of Technology in the turn of the millenium. The mean data over all listeners and test sequences are shown in Fig. 15 of [P8]. The results indicate that WLPC is superior to LPC especially at high sampling rates and for low orders of filters. An early version of this article was published in (Härmä 2000).

3.4 Low-delay audio coding

As was discussed in Section 2.3.2, parametric coding techniques can be modified in such a way that spectral modeling is completely or partially based on already transmitted signal. With LPC, this is called *backward adaptive predictive coding* (Gibson et al. 1974). In [P2], an extremely low delay audio codec was proposed. This codec, which is actually a warped implementation of an algorithm proposed in (Iyengar & Kabal 1988), has the coding delay of 1-2 samples. The spectral modeling is based on *Gradient Adaptive Lattice*, GAL, method (Yatrou & Mermelstein 1988, Iyengar & Kabal 1988)[17] which is driven by the *decoded* signal. The codec is a stereo codec with a common vector quantizer [18] At 220 kb/s for stereo the codec performed surprisingly well with smoothly varying audio test material. However, in the case of sudden transients, the output was a disaster. This can be expected because the codec has no means to adapt to a sudden onset of a signal.

[15] Especially in the case of IIR Hilbert transformers.

[16] Quantization of filter coefficients was largely omitted in this paper. An article about this topic was published in (Vaalgamaa, Härmä & Laine 2000). Warped filters are generally known to be relatively robust for quantization of filter coefficients, see, e.g., (Asavathiratham et al. 1999).

[17] GAL was also used by Fejzo & Lev-Ari (1997) for adaptive Laguerre filtering.

[18] An early version of this algorithm was actually implemented as a single complex-valued process as in [P1]. This works, but the final coding delay is increased by tens of milliseconds due to the mappings from stereo signal pair to a complex-valued signal and back.

The work with low-delay coding continued in two conference articles: [P3] and [P4]. Both of these papers mainly concentrate on the theory of perceptual low-delay wideband audio coding. It turned out that relatively little has been done to answer basic questions such as what is a sufficiently low coding delay and what type of techniques are available for the development of such coders.

It cannot be avoided that in low-delay coding the estimated spectral model is inaccurate for sudden onsets and transients. There is simply no time or signal data available for a detailed analysis. In [P3], it was assumed that the same problem could also be found in human perception, that is, the ear cannot accurately detect spectral details during and immediately after an onset or a transient. It was proposed that this could be related to a well-known psychoacoustic phenomenon, namely *overshoot masking*, which reduced the sensitivity of the ear immediately after the onset of a wideband sound. The article reviews some results from psychoacoustics and introduces a simulated low-delay codec which was used in listening tests to test this idea. It turned out that a significant amount of quantization noise can be tolerated immediately after the onset of a wideband sound or a transient.

The main result of [P3] is that low-delay high-quality audio coding is possible even if performance is necessarily degraded near sudden onsets and transients. However, the author was not able to find a computational auditory model for the overshoot effect which could accurately predict the results. One such model was developed in (Härmä 1999). However, it is computationally too complex to be applied to any practical audio or speech coding algorithm.

The simulated codec introduced in [P3] was based on a *warped sliding-window lattice method*. The method was studied earlier in (Härmä 1998*a*) and it is close to some of the techniques presented, e.g., in (Zhao, Ling, Lev-Ari & Proakis 1994, Demeure & Scharf 1990).

Another low-delay wideband audio codec was introduced in [P4]. This can be seen as a modified version of G.728 speech coder (Chen et al. 1992). The conventional Low-Delay CELP algorithm was warped and applied to wideband audio. In addition, several additional techniques were presented in the article. However, the algorithm was never fully implemented and tested. The main reason is that the presented algorithm is clearly suboptimal. It is based on backward adaptive warped LP with a 5 ms look-ahead buffer, which is used very inefficiently in the codec. The use of backward adaptive LPC is well justified in low-delay coding, but it is difficult to build an efficient coding algorithm which uses both backward and *forward* adaptive spectral modeling. This calls for new techniques. The sliding-window lattice method which was used in simulated low-delay coding in [P3], would be an attractive technique if an efficient method for parametrization was found.

The main contribution of this article [P4] is an extensive discussion on the requirements for algorithmic coding delay in various applications based on bidirectional audio transmission, such as conventional teleconferencing and teleimmersive virtual reality applications (Zyda 1992, Huopaniemi 1999, Savioja 1999). It was estimated that the coding delay in the range of 2 to 10 milliseconds should be sufficient for most of audio applications.

3.5 Time-varying spectral modeling

Natural audio and speech signals are nonstationary processes. The assumption of local stationarity works relatively well in many applications. However, in many audio and speech coding techniques, including those presented in this thesis, difficulties are encountered in encoding transients, onsets, and rapid chirp-like signals efficiently. This is a significant problem especially in low-delay coding because it is not possible to use long buffering and associated *bit-resevoir* techniques.

In [P2], [P3], and [P4] is was shown that warped signal processing techniques can be applied to certain adaptive, and sliding-window, formulations of linear predictive spectral modeling. Adaptive LMS formulations for Laguerre filters had been already introduced in (den Brinker 1993, den Brinker 1994), Laguerre GAL algorithm in (Fejzo & Lev-Ari 1997), and RLS Laguerre lattice algorithm in (Merced & Sayed 2000). In [P5], it is demonstrated how deterministic regression time-varying LPC techniques, see Sect. 2.3.3, can be warped. This technique yields an efficient, and perceptually motivated parametrization for time-varying sounds. However, the direct-form time-varying autoregressive method (Subba Rao 1970, Liporace 1975) has some problems. Therefore, the recent work by the current author involves the utilization of a modified[19] version of a time-varying lattice method introduced by Grenier (1983).

3.6 Future work

Articles in this thesis introduce and study a large number of techniques for which frequency-warping techniques can be applied to. Potential applications are also discussed. However, no fully tested and tuned applications have been presented so far. This is something that can be expected to take place in the future. One of the main thesis of this work is that warping techniques can be applied to basically any DSP algorithm. Therefore, warped linear prediction can be used in audio and speech codecs as a replacement for the classical LP. It is also stated and demonstrated that it is reasonable to assume that the use of WLP may lead to subjectively better performance of a codec due to a better match with the properties of human hearing.

The results in [P8] show that significant saving in quantization of residual signal can be obtained by using warping techniques in LPC. In another article (Vaalgamaa et al. 2000), it was shown that the quantization properties of warped or conventional filter parameters are nearly equal. However, it turned out that slightly more bits were typically needed to quantize WLP coefficients than conventional LPC parameters. The benefits of using WLP in audio and speech coding can be evaluated only by designing a fully optimized codec.

[19] That is, a warped version.

Subband coders and different classes of sinusoidal+harmonics+noise coders are powerful tools for audio bitrate reduction. However, they typically suffer from a relative high algorithmic coding delay. In [P4] it was proposed that the use of parametric techniques, and WLP, in particular, would be beneficial in low-delay codecs. At the moment, the main application field of the author is wideband telecommunications. This involves development of high-quality, multi-channel, and low-delay audio coding techniques, and integration of those with other elements of such systems, i.e., acoustic echo cancellation and channel coding.

Many parts of this thesis open perspectives which may be subjects to the future work of the author. Generalized predictor structures studied in [P7] is clearly a field where more work both in theoretical aspects and practical applications could be done. The work with time-varying autoregressive techniques started with [P5] is also continuing. The main problem with both these themes is in finding the best applications. Generalized models, time-varying models, and generalized time-varying models might be used in many types of applications based on analysis or synthesis of audio or speech signals.

4. Conclusions

4.1 Main results of this thesis

- Implementation techniques for frequency-warped and generalized recursive filters have been introduced in [P6]. Two techniques have been presented. First, a generic algorithm to implement recursive filters which are traditionally considered to be *non-realizable* due to delay-free loops. Secondly, technique to derive a corresponding modified filter structure, where the delay-free loops are eliminated.
- A generalization and a set of new modifications to the predictive coding have been introduced in [P7].
- Most of conventional signal processing techniques can be warped. This topic has been reviewed and extended in [P9]. This article also introduced a new design of a warped IIR filterbank and some new approaches for filter design.
- In [P8], the performance of warped linear prediction in audio coding have been studied in terms of technical measures and listening tests at different sampling rates and as a function of model order.
- A new formulation for stereo audio coding has been presented in [P1].
- Several new wideband audio coding algorithms have been introduced in [P1], [P2], and [P4]. In the context of this thesis, these can be considered as design examples.
- Psychoacoustical, acoustical, and technical aspects of low-delay wideband audio coding have been studied in [P3] and [P4]. It is proposed that warped linear prediction may be a potential technique for low-delay wideband audio coding.
- A warped formulation of time-varying autoregressive modeling has been presented in [P5] and its applicability to the modeling of audio and speech signals has been studied.

4.2 Contribution of the author

The author of this thesis produced approximately 68 % of the pages of the manuscript [P9]. It was accepted by the other authors of the article that the current author is responsible for approximately 80 % of the technical work which includes mathematics, computer simulations, and programs made for this particular article. It was also desided that the total contribution of the current author for this article is 75 %. Naturally, if the contribution of the author in both writing and techical part would be 100 %, the total contribution would also be 100 %. Therefore, we have here a system of two equations where one can easily solve weights for written and technical parts. This gives weights 5/12 and 7/12 for writing and technical contribution, respectively. This formula was used in calculating the total contribution of the author of this thesis for all the articles (marked in parenthesis). The percentage values for writing and technical contribution was evaluated together with other authors and acknowledged contributors.

P1 The principle of complex-valued warped LPC was developed together with Laine. The author wrote the article and did all computational simulations. (83 %)

P2 The author wrote the article, developed the coding algorithm, and designed and conducted listening test. Co-authors Laine and Karjalainen helped to enhance the quality of the final article. (95 %)

P3 The author wrote the article, developed the presented technique and designed and conducted listening tests. Co-authors Laine and Karjalainen helped to enhance the quality of the final article. (94 %)

P4 The author wrote the article, developed the presented coding algorithm. Co-authors Laine and Karjalainen helped to enhance the quality of the final article. (95 %)

P5 The topic of this article was developed together with Juntunen. He also introduced the basic methodology to the author and wrote the original version of Section 2. The current author *warped* the methodology and did all computational simulations. Juntunen helped in enhancing the quality and language of the final article. (82 %)

P6 The implementation techniques were developed by the current author and he also wrote the article. Several people are acknowledged for help in formulating the final article. (97 %)

P7 This work was partly inspired by earlier work by Laine. The author wrote the article and did all computational simulations. Several people helped the author by providing useful Matlab code fragments for filter design, and in critical reading of an early version of the manuscript. (96 %)

P8 The author wrote the article and developed the methodology for the tests. He also designed the listening test system and conducted the listening tests. Laine and Alku read an early version of the manuscript and helped in developing the quality of the presentation. (95 %)

P9 The idea for the article came from Karjalainen, who also wrote the original version of *Introduction* and Section 2.5. The co-authors provided specific application examples, i.e., Sections 3.4 – 3.7. All other material was written and computed by the current author. The role of Karjalainen and Välimäki in enhancing the quality and language of the final manuscript was vital. (75 %)

Bibliography

Asavathiratham, C., Beckmann, P. E. & Oppenheim, A. V. (1999), Frequency warping in the design and implementation of fixed-point audio equalizers, *in* ‘Proc. IEEE Workshop Appl. Signal Proc. Audio and Acoust.’, IEEE, New Paltz, New York, pp. 55–59.

Åström, K. J. (1970), *Introduction to Stochastic Control Theory*, Academic Press Inc., New York.

Atal, B. (1982), ‘Predictive coding of speech at low bit rates’, *IEEE Trans. Comm.* **COM-30**(4), 600–614.

Atal, B. & Hanauer, S. L. (1971), ‘Speech analysis and synthesis by linear prediction of the speech wave’, *J. Audio Eng. Soc.* **50**, 637–655.

Atal, B. S. & Schroeder, M. R. (1967), Predictive coding of speech signals, *in* ‘Proc. 1967 IEEE Conf. on Communication and Processing’, pp. 360–361.

Barnwell, T. (1977), Recursive autocorrelation computation for LPC analysis, *in* ‘Proc. of IEEE Int. Conf. on Acouctics, Speech, and Signal Processing’, Hartford, pp. 1–4.

Beerends, J. G. & Stemerdink, J. A. (1996), ‘A perceptual audio quality measure based on a psychoacoustic sound representation’, *J. Audio Eng. Soc.* **40**, 963–978.

Berger, T. & Gibson, J. D. (1998), ‘Lossy source coding’, *IEEE Trans. Inform. Theory* **44**(6), 2693–2723.

Bessette, B., Salami, R., Laflamme, C. & Lefebvre, R. (1999), A wideband speech and audio codec at 16/24/32 kbit/s using hybrid ACELP/TCX techniques, *in* ‘Proc. IEEE Workshop on Speech Coding’, IEEE, Porvoo, Finland.

Blackman, R. B. & Tukey, J. W. (1959), *the Measurement of Power Spectra*, Dover, New York.

Boland, S. & Deriche, M. (1995), high quality audio coding using multipulse *lpc* and wavelet decomposition, *in* ‘Proc. Int. Conf. Acoust., Speech, and Signal Proc.’, IEEE, Detroit, USA, pp. 3067–3069.

Booton, R. C. (1952), ‘An optimization theory for time-varying linear systems with nonstationary inputs’, *Proc. IRE* **40**, 977–981. Cited from (Kalman 1960).

Box, G. E. & Jenkins, G. M. (1970), *Time Series Analysis, Forecasting and Control*, Holden-Day. Cited from (Makhoul & Cosell 1976).

Braccini, C. & Oppenheim, A. V. (1974), ‘Unequal bandwidth spectral analysis using digital frequency warping’, *IEEE Trans. Acoust., Speech, and Signal Processing* **ASSP-22**(4), 233–245.

Brandenburg, K. (1994), ‘ISO-MPEG-1 audio: A generic standard for coding of high-quality digital audio’, *J. Audio Eng. Soc.* **42**, 780–792.

Brandenburg, K. (1998), Perceptual coding of high quality digital audio, *in* M. Kahrs & K. Brandenburg, eds, ‘Applications of Digital Signal Processing to Audio and Acoustics’, Kluwer Academic Press, pp. 39–83.

Brandenburg, K. & Sporer, T. (1992), “NMR” and “masking flag”: Evaluation of quality using perceptual criteria, *in* ‘Proc. 11th Int. AES Conf’, Portland.

Brandenburg, K., Langenbucher, G. G., Schramm, H. & Seitzer, D. (1982), A digital signal processor for real time adaptive transform coding of audio signal up to 20 khz bandwidth, *in* ‘Proc. ICCC’, pp. 474–477.

Burg, J. (1975), Maximum entropy spectral analysis, PhD thesis, Stanford University, Stanford.

Burg, J. P. (1967), Maximum entropy spectral analysis, *in* ‘Proc. 37th Meeting Soc. of Exploration Geophysicists’, Oklahoma City, Oklahoma.

Chang, K. F., Cheong, P., Ting, S. W. & Tam, K. W. (2000), Novel speech all-pole modeling based upon selective even-samples linear prediction, *in* ‘Proc. IEEE Int. Conf. Electronics, Circuits and Systems’, pp. 1022–1025.

Chen, J.-H. (1995), *Speech Synthesis and Coding*, Elsevier Science Publ., Amsterdam, the Netherlands, chapter Low-Delay Coding of Speech, pp. 209–256. In (Kleijn & Paliwal 1995*c*).

Chen, J.-H. & Cox, R. V. (1993), 'The creation and evolution of 16 kb/s LD-CELP: from concept to standard', *Speech Communication* **12**(2), 103–111. A special issue on G.728 coder.

Chen, J.-H. & Wang, D. (1996), Transform predictive coding of wideband speech signals, *in* 'Proc. of IEEE Int. Conf. on Acouctics, Speech, and Signal Processing', Vol. I, Atlanta, USA, pp. 275–278.

Chen, J.-H., Cox, R. V., Lin, Y.-C. & Jayant, N. (1992), 'A low-delay CELP coder for the CCITT 16 kb/s speech coding standard', *IEEE J. Sel. Areas in Comm.* **10**(5), 830–849.

Chen, J.-H., Lin, Y.-C. & Cox, R. V. (1991), A fixed-point 16 kb/s LD-CELP algorithm, *in* 'Proc. of IEEE Int. Conf. on Acouctics, Speech, and Signal Processing', IEEE, pp. 21–24.

Constantinides, A. G. (1970), 'Spectral transformations for digital filters', *Proc. IEEE* **117**(8), 1585–1590.

Cooley, J. W. & Tukey, J. W. (1965), 'An algorithm for machine computation of complex fourier series', *Math. Comp.* **19**(90), 297–301. Reprinted in (Rabiner & Rader 1972).

Cremer, H. (1961), On some classes of nonstationary processes, *in* 'Proc. 4th Berkeley Symp. Math., Statist., Probability', Vol. 2, Univ. California Press, Berkeley, CA, USA, pp. 57–78.

Crochiere, R. E., Webber, S. M. & Flanagan, J. K. L. (1976), 'Digital coding of speech in sub-bands', *Bell Syst. tech. J.* **55**, 1069–1086.

Cutler, C. C. (1952), 'Differential quantization of communications', U.S. Patent 2 605 361. Cited from (Berger & Gibson 1998).

Delsarte, P. & Genin, Y. (1986), 'The split Levinson algorithm', *IEEE Trans. Acoust., Speech and Signal Processing* **34**, 470–478.

Demeure, C. J. & Scharf, L. L. (1990), 'Sliding windows and lattice algorithms for computing QR factors in the least squares theory of linear prediction', *IEEE Trans. Acoust., Speech, and Signal Proc.* **38**(4), 721–725.

den Brinker, A. C. (1993), 'Adaptive modified Laguerre filters', *Signal Processing* **31**(1), 69–79.

den Brinker, A. C. (1994), 'Laguerre-domain adaptive filters', *IEEE Trans. Signal Processing* **42**(4), 953–956.

den Brinker, A. C. (1998), The auditory critical bands interpreted as a local Kautz transformation, *in* 'Signal Processing IX: Theories and Applications', Vol. I, EURASIP, Rhodes, Greece, pp. 125–128.

Denes, P. & Pinson, E. (1963), *The Speech Chain*, Anchor books, New York, USA. Cited from (Kleijn & Paliwal 1995*a*).

Dimino, G. & Morpurgo, L. (1996), Intra-channel prediction: a new tool for the subband coding of high quality audio, *in* 'AES 100st Convention preprint 4198', AES, Copenhagen, Denmark.

Doob, J. L. (1944), 'The elementary Gaussian processes', *Ann. Math. Statist.* **15**, 229–282.

Durbin, J. (1960), 'The fitting of time-series models', *Rev. Inst. Int. Statist.* **28**(3), 233–243.

Edler, B. & Schuller, G. (2000), Audio coding using a psychoacoustic pre- and post-filter, *in* 'Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.', Vol. II, IEEE, Istanbul, Turkey, pp. 881–884.

Elias, P. (1955), 'Predictive coding, Parts I and II', *IRE Trans. Inform. Theory* **IT-1**, 16–33. Cited from (Makhoul 1975).

Evangelista, G. & Cavaliere, S. (1998), 'Discrete frequency warped wavelets: Theory and applications', *IEEE Trans. Signal Processing* **46**(4), 874–885.

Fejzo, Z. & Lev-Ari, H. (1997), 'Adaptive Laguerre-lattice filters', *IEEE Trans. Signal Processing* **45**(12), 3006–3016.

Fielder, L. D., Bosi, M., Davidson, G., Davis, M., Todd, C. & Vernon, S. (1996), AC-2 and AC-3: Low-complexity transform-based audio coding, *in* N. Gilchrist & C. Grewin, eds, 'Collected Papers on Digital Audio Bit-Rate Reduction', Audio Engineering Society Inc., pp. 54–72.

Fletcher, H. (1953), *Speech and Hearing in Communications*, Van Nostrand, Princeton, USA.

Friedlander, B. (1982), 'Lattice methods of spectral estimation', *Proc. IEEE* **70**(9), 991–1016.

Gersho, A. (1994), 'Advances in speech and audio compression', *Proc. IEEE* **82**(6), 900–918.

Gerzon, M. A., Graven, P. G., Stuart, J. R., Law, M. J. & Wilson, R. J. (1999), The MLP lossless compression system, *in* 'Proc. AES 17th Int. Conf.: High-Quality Audio Coding', AES, Florence, Italy, pp. 61–75.

Gibson, J. (1980), 'Adaptive prediction in speech differential encoding systems', *Proc. IEEE* **68**(4), 488–525.

Gibson, J. D., Cheong, Y. C., Chang, W.-W. & Woo, H. C. (1990), A comparison of backward adaptive prediction algorithms in low delay speech coders, *in* 'Proc. of ICASSP'90', Vol. 1, IEEE, Albuquerque, New Mexico, pp. 237–240.

Gibson, J. D., Jones, S. K. & Melsa, J. L. (1974), 'Sequentially adaptive prediction and coding of speech signals', *IEEE Trans. on Comm.* **22**(11), 1789–1797.

Gottlieb, M. J. (1938), ‘Polynomials orthogonal on a finite or numerable set of points’, *Am. J. Math.* **60**, 453–458.

Gray, R. M. & Neuhoff, D. L. (1998), ‘Quantization’, *IEEE Trans. Inform. Theory* **44**(6), 2325–2383.

Grenier, Y. (1983), ‘Time-dependent ARMA modeling of nonstationary signals’, *IEEE Trans. Acoust. Speech and Signal Proc.* **ASSP-31**, 899–911.

Griffiths, L. J. (1977), A continuously-adaptive filter implemented as a lattice structure, *in* ‘Proc. Int. Conf. Acoust. Speech, and Signal Proc.’, IEEE, Hartford, USA, pp. 683–686.

Grill, B. (1999), The MPEG-4 general audio coder, *in* ‘Proc. AES 17th Int. Conf.: High-Quality Audio Coding’, AES, Florence, Italy, pp. 147–156.

Hall, M., Oppenheim, A. & Willsky, A. (1983), ‘Time-varying parametric modeling of speech’, *Signal Processing* **5**, 267–285.

Hamdy, K. N., Ali, M. & Tewfik, A. H. (1996), Low bit rate high quality audio coding with combined harmonic and wavelet representations, *in* ‘Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.’, Vol. 2, IEEE, Atlanta, USA, pp. 1045–1048.

Hardwick, J. C. & Lim, J. S. (1988), A 4.8 bps improved multi-band excitation speech coder, *in* ‘Proc. of IEEE Int. Conf. on Acouctics, Speech, and Signal Processing’, pp. 374–377.

Haykin, S. (1989), *Modern Filters*, Macmillan, New York.

Haykin, S. (1996), *Adaptive Filter Theory*, 3 edn, Prentice-Hall, Inc., New Jersey.

Hedelin, P. (1981), A tone-oriented voice-excited vocoder, *in* ‘Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.’, Vol. I, IEEE, Atlanta, USA, pp. 205–208.

Hermansky, H. (1990), ‘Perceptual linear predictive (plp) analysis of speech’, *J. Acoust. Soc. Am.* **87**(4), 1738–1752.

Härmä, A. (1997), Perceptual aspects and warped techniques in audio coding, Master’s thesis, Helsinki University of Technology, Espoo, Finland. p. 88.

Härmä, A. (1998*a*), Audio coding with warped predictive methods, Licentiate’s Thesis, Helsinki University of Technology, Espoo, Finland. p. 104.

Härmä, A. (1998*b*), Implementation of recursive filters having delay free loops, *in* ‘Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing’, Vol. III, Seattle, Washington, pp. 1261–1264.

Härmä, A. (1999), Low-level auditory modeling of temporal effects, *in* H. G. Okuno, ed., ‘16th Int. Joint Conf. on Artificial Intelligence, Workshop on Computational Auditory Scene Analysis’, Stockholm, Sweden, pp. 1–9.

Härmä, A. (2000), Evaluation of a warped linear predictive coding scheme, *in* ‘Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing’, Vol. II, IEEE, Istanbul, Turkey, pp. 897–900.

Härmä, A., Laine, U. K. & Karjalainen, M. (1996), Warped linear prediction in audio coding, *in* ‘Proc. IEEE Nordic Signal Proc. Symposium, NORSIG’96’, Espoo, Finland, pp. 447–450.

Härmä, A., Laine, U. K. & Karjalainen, M. (1997), WLPAC – a perceptual audio codec in a nutshell, *in* ‘AES 102nd Conv. preprint 4420’, Munich, Germany.

Härmä, A., Vaalgamaa, M. & Laine, U. K. (1998), A warped linear predictive stereo codec using temporal noise shaping, *in* ‘Proc. Nordic Signal Proc. Symposium, NORSIG’98’, Denmark, pp. 229–232.

Huopaniemi, J. (1999), Virtual Acoustics and 3-D Sound in Multimedia Signal Processing, PhD thesis, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, Espoo, Finland. Report no. 53.

Huopaniemi, J., Zacharov, N. & Karjalainen, M. (1999), ‘Objective and subjective evaluation of head-related transfer function filter design’, *J. Audio Eng. Soc.* **47**(4), 218–239.

Imai, S. (1980), ‘Log magnitude approximation (LMA) filter (in japanese)’, *Trans. IECEJ* **63-A**(12), 886–893.

Imai, S. (1983), Cepstral analysis synthesis on the mel frequency scale, *in* ‘Proc. ICASSP’83’, IEEE, Boston.

Imai, S. & Furuichi, C. (1988), Unbiased estimator of log spectrum and its application to speech signal processing, *in* ‘Signal Processing IV:Theories and Applications’, Vol. 1, EURASIP, Elsevier, pp. 203–206.

ISO/IEC (1993), 11172-3, information technology-coding of part 3: Audio, Technical report, ISO/IEC. Standard for MPEG-1 Audio.

Itakura, F. (1975), ‘Line spectrum representation of linear predictive coefficients of speech signals’, *J. Acoust. Soc. Am.* **57**, S35.

Itakura, F. & Saito, S. (1970), ‘A statistical method for estimation of speech spectral density and formant frequencies’, *Electr. and Commun. in Japan* **52-A**, 36–43.

Itakura, F. & Saito, S. (1972), On the optimum quantization of feature parameters in the PARCOR speech synthesizer, *in* ‘Proc. Conf. Speech Commun. and Processing’, pp. 434–437. Reprinted in (Schafer & Markel 1979).

Iwakami, N. & Moriya, T. (1996), Transform-domain weighted interleave vector quantization, *in* 'AES 101st Convention preprint 4377', AES, Los Angeles, USA.

Iyengar, V. & Kabal, P. (1988), A low delay 16 kbits/sec speech coder, *in* 'Proc. of ICASSP'88', Vol. 1, IEEE, New York, pp. 243–246.

Jayant, N. S. (1973), 'Adaptive quantization with one word memory', *Bell System Tech. Journal* pp. 1119–1144.

Jayant, N. S. & Noll, P. (1984), *Digital coding of waveforms*, Prentice-Hall, New Jersey.

Jayant, N. S., Johnston, J. D. & Sefranek, R. (1993), 'Signal compression based on models of human perception', *Proc. IEEE* **81**(10), 1385–1422.

Johnson, D. H. (1976), Application of digital-frequency warping to recursive variable-cutoff digital filters, *in* 'Journal of the Illuminating Engineering Society IEEE Electron and Aerosp Syst Conv (EASCON '76)'.

Johnston, J. D. (1988), Estimation of perceptual entropy using noise masking criteria, *in* 'Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.', IEEE, New York, USA, pp. 2524–2527.

Johnston, J. D. & Brandenburg, K. (1992), Wideband coding – perceptual considerations for speech and music, *in* S. Furui & M. M. Sondhi, eds, 'Advances in Speech Signal Processing', Marcel-Dekker, chapter 4, pp. 109–140.

Johnston, J. D., Sinha, D., Dorward, S. & Quackenbush, S. (1996), AT&T perceptual audio coding (PAC), *in* N. Gilchrist & C. Grewin, eds, 'Collected Papers on Digital Audio Bit-Rate Reduction', Audio Engineering Society Inc., pp. 73–82.

Juntunen, M. (1999), Direct Stabilization of Autoregressive Models, PhD thesis, University of Kuopio, Kuopio, Finland.

Kailath, T. (1974), 'A view of three decades of linear filtering theory', *Trans. Information Theory* **20**(2), 146–181. Reprinted in (Kailath 1977).

Kailath, T., ed. (1977), *Linear Least.Squares Estimation*, Vol. 17 of *Benchmark Papers in Electr. Eng. and Computer Science*, Dowden, Hutchinson & Ross Inc., Pennsylvania.

Kaipio, J. P. & Juntunen, M. (1999), Deterministic regression smoothness priors tvar modelling, *in* 'Proc. IEEE Int. Conf. Acoust. Speech, and Signal Proc.', Vol. III, IEEE, Phoenix, Arizona, USA, pp. 1693–1696.

Kalman, R. E. (1960), 'A new approach to linear filtering and prediction problems', *J. Basic Eng.* **83**, 34–45.

Karjalainen, M. (1985), A new auditory model for the evaluation of sound quality of audio system, *in* 'Proc. of ICASSP'85', IEEE, Florida, pp. 608–611.

Karjalainen, M. & Smith, J. O. (1996), Body modeling techniques for string instrument synthesis, *in* 'Proc. Int. Comp. Music Conf.', Hong Kong, pp. 232–239.

Karjalainen, M., Härmä, A. & Laine, U. K. (1996), Realizable warped IIR filter structures, *in* 'Proc. of the IEEE Nordic Signal Proc. Symposium, NORSIG 96', Espoo, Finland, pp. 483–486.

Karjalainen, M., Härmä, A. & Laine, U. K. (1997), Realizable warped IIR filters and their properties, *in* 'Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing', Vol. 3, Munich, pp. 2205–2209.

Karjalainen, M., Piirilä, E., Järvinen, A. & Huopaniemi, J. (1999), 'Comparison of loudspeaker equalization methods based on DSP techniques', *J. Audio Eng. Soc.* **47**(1/2), 15–31.

Kautz, W. H. (1954), 'Transient synthesis in the time domains', *IRE Trans. Circuit Theory* **CT-1**(3), 29–39.

Khintchine, A. (1934), 'Korrelationstheorie der stationären stochastischen Prozesse', *Mathem. Ann.* Cited from (Wold, 1938).

King, R. E. & Paraskevopoulos, P. N. (1977), 'Digital Laguerre filters', *J. Circuit Theory Applicat.* **5**, 81–91.

Kleijn, W. B. & Paliwal, K. K. (1995*a*), *Speech Synthesis and Coding*, Elsevier Science Publ., Amsterdam, the Netherlands, chapter An introduction to speech coding, pp. 209–256. In (Kleijn & Paliwal 1995*c*).

Kleijn, W. B. & Paliwal, K. K. (1995*b*), *Speech Synthesis and Coding*, Elsevier Science Publ., Amsterdam, the Netherlands, chapter Waveform interpolation for coding and synthesis, pp. 175–207. In (Kleijn & Paliwal 1995*c*).

Kleijn, W. B. & Paliwal, K. K., eds (1995*c*), *Speech Synthesis and Coding*, Elsevier Science Publ.

Kobayashi, T. & Imai, S. (1984), 'Spectral analysis using generalized cepstrum', *IEEE Trans. Acoust. Speech, and Signal Proc.* **ASSP-32**(5), 1087–1089.

Koishida, K., Hirabayashi, G., Tokuda, K. & Kobayashi, T. (1998), A wideband CELP speech coder at 16 kbit/s based on mel-generalized cepstral analysis, *in* 'Proc. of ICASSP'98', Vol. 1, IEEE, Seattle, pp. 161–164.

Koishida, K., Tokuda, K., Kobayashi, T. & Imai, S. (1996), CELP coding system based on mel-generalized cepstral analysis, *in* 'Proc. Int. Conf. Spoken Lang. Proc.', Vol. 1, Philadelphia.

Kolmogorov, A. N. (1941), 'Stationary sequences in Hilbert space', *Bull. Math. Univ. Moscow*. English translation available in (Kailath 1977).

Kroon, P. & Kleijn, W. B. (1995), *Speech Synthesis and Coding*, Elsevier Science Publ., Amsterdam, the Netherlands, chapter Linear-Prediction based Analysis-by-Synthesis Coding, pp. 79–119. In (Kleijn & Paliwal 1995*c*).

Krüger, E. & Strube, H. W. (1988), 'Linear prediction on a warped frequency scale', *IEEE Trans. Acoust. Speech, and Signal Proc.* **36**(9), 1529–1531.

Laakso, T. I., Välimäki, V., Karjalainen, M. & Laine, U. K. (1996), 'Splitting the unit delay – tools for fractional delay filter design', *IEEE Signal Processing Magazine* pp. 30–60.

Laine, U. K. (1992), Famlet, to be or not to be a wavelet, *in* 'Proc. Int. Symp. Time-Frequency and Time-Scale Analysis', IEEE, Victoria, Canada, pp. 335–338.

Laine, U. K. (1995), Generalized linear prediction based on analytic signals, *in* 'Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.', Vol. III, IEEE, Detroit, Michican, USA, pp. 1701–1704.

Laine, U. K. & Härmä, A. (1996), Bark-famlet filterbanks, *in* 'Proc. Nordic Acoustical Meeting', Helsinki, Finland, pp. 277–284.

Laine, U. K., Karjalainen, M. & Altosaar, T. (1994), WLP in speech and audio processing, *in* 'Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.', Vol. III, Adelaide, pp. 349–352.

Lee, Y. W. (1933), 'Synthesis of electric networks by means of the fourier transforms of Laguerre's functions', *J. Mathematics and Physics* **XI**, 83–113.

Lee, Y. W. (1960), *Statistical theory of communication*, Wiley, New York.

Lefebvre, R., Salami, R., Laflamme, C. & Adoul, J.-P. (1993), 8 kbits/s coding of speech with 6 ms frame-length, *in* 'Proc. of IEEE Int. Conf. on Acouctics, Speech, and Signal Processing', Minneapolis,USA, pp. 612–615.

Levine, S. N. & Smith, J. O. (1999), A switched parametric & transform audio coder, *in* 'Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.', Vol. II, Phoenix, USA, pp. 985–988.

Levinson, N. (1947), 'The Wiener RMS (root mean square) error criterion in filter design and prediction', *J. Math. Phys.* **25**(4), 261–278. Also an appendix in (Wiener 1949).

Li, G. (1998), Designing low-pass digital filters with fewer parameters, *in* 'Signal Processing IX: Theories and Applications', Vol. III, EURASIP, Rhodes, Greece, pp. 1901–1904.

Lin, S. & Costello Jr., D. J. (1983), *Error control coding: fundamentals and applications*, Prentice-Hall, New Jersey, USA.

Lin, X. & Steele, R. (1993), Subband coding with modified multipulse LPC for high quality audio, *in* 'Proc. Int. Conf. Acoust., Speech, and Signal Proc.', IEEE, Minneapolis, USA, pp. 201–204.

Liporace, L. (1975), 'Linear estimation of nonstationary signals', *J. Acoust. Soc. Am.*

Maione, B. & Turchiano, B. (1985), 'Laguerre z-transfer function representation of linear discrete-time systems', *Int. J. of Control* **41**(1), 245–257.

Makhoul, J. (1975), 'Linear prediction: A tutorial review', *Proc. IEEE* pp. 561–580. Reprinted in (Schafer & Markel 1979).

Makhoul, J. (1977), 'Stable and efficient lattice methods for linear prediction', *IEEE Trans. Acoust., Speech, and Signal Proc.* **ASSP-25**(5), 423–428.

Makhoul, J. & Berouti, M. (1979), 'Adaptive noise spectral shaping and entropy coding in predictive coding of speech', *IEEE Trans. Acoust., Speech, and Signal Proc.* **ASSP-27**(1), 63–73.

Makhoul, J. & Cosell, L. (1976), LPCW:an lpc coder with linear predictive spectral warping, *in* 'Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.', Philadelphia, pp. 466–469.

Markel, J. D. & Gray, A. H. (1976), *Linear Prediction of Speech*, Vol. 12 of *Communication and Cybernetics*, Springer-Verlag, New York.

McAulay, R. J. & Quatieri, T. F. (1986), 'Speech analysis/synthesis based on a sinusoidal speech model', *IEEE Trans. Acoust., Speech, and Signal Proc.* **34**, 744–754.

McCree, A. V. & Barnwell III, T. P. (1995), 'A mixed excitation LPC vocoder model for low bit rate speech coding', *IEEE Trans. Speech and Audio Processing* **3**(4), 242–250.

Merced, R. & Sayed, A. H. (2000), Exact RLS Laguerre-lattice adaptive filtering, *in* 'Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.', Vol. I, IEEE, Istanbul, Turkey, pp. 456–459.

Messerschmitt, D. G. (1980), 'A class of generalized lattice filters', *IEEE Trans. Acoust., Speech, and Signal Proc.* **ASSP-28**(2), 198–204.

Moore, B. C. J. (1997), *Introduction to the psychology of hearing*, 4th edition edn, Academic Press.

Moore, B. C. J., Oldfield, S. D. & Dooley, G. J. (1989), ‘Detection and discrimination of spectral peaks and notches at 1 and 8 khz’, *J. Acoust. Soc. Am.* **85**(2), 820–836.

Moore, B. C. J., Peters, R. W. & Glasberg, B. R. (1990), ‘Auditory filter shapes at low center frequencies’, *J. Acoust. Soc. Am.* **88**(1), 132–140.

Moorer, J. A. (1979), ‘The digital coding of high-quality musical sound’, *J. Audio Eng. Soc.* **27**(9), 657–666.

Moriya, T., Iwakami, N., Ikeda, K. & Miki, S. (1996), Extension and complexity reduction of TwinVQ audio coder, *in* ‘Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.’, Vol. 2, IEEE, Atlanta, USA, pp. 1029–1032.

Muraoka, T., Iwahara, M. & Yamada, Y. (1981), ‘Examination of audio-bandwidth requirements for optimum sosuns signal transmission’, *J. Audio Eng. Soc.* **29**(1/2), 2–9.

Ninness, B. & Gustafsson, F. (1997), ‘A unifying construction of orthogonal bases for system identification’, *IEEE Trans. Automatic Control* **42**(4), 515–521.

Nishiguchi, M. (1999), MPEG-4 speech coding, *in* ‘Proc. AES 17th Int. Conf.: High-Quality Audio Coding’, AES, Florence, Italy, pp. 139–146.

Noll, P. (1975), ‘A comparative study of various quantization schemes for speech encoding’, *Bell System Tech. Journal* **54**(9), 1597–1614.

Noll, P. (1978), ‘On predictive quantizing schemes’, *Bell System Tech. Journal* **57**(5), 1499–1532.

Oliveira e Silva, T. (1995*a*), ‘Laguerre filters – an introduction’, *Revista do Detua* **1**(3), 237–248.

Oliveira e Silva, T. (1995*b*), ‘Optimality conditions for truncated Laguerre networks’, *IEEE Trans. Signal Processing* **42**(9), 2528–2530.

Oliver, B. M. (1952), ‘Efficient coding’, *Bell System Tech. Journal* **31**, 724–750.

Oppenheim, A. V. & Johnson, D. H. (1972), ‘Discrete representation of signals’, *Proc. IEEE* **60**(6), 681–691.

Oppenheim, A. V. & Schafer, R. W. (1968), ‘Homomorphic analysis of speech’, *Trans. Audio and Electro-acoustics* **AU-16**, 221–226.

Oppenheim, A. V., Johnson, D. H. & Steiglitz, K. (1971), ‘Computation of spectra with unequal resolution using the Fast Fourier Transform’, *Proc. IEEE* **59**, 299–301.

Paatero, T. (2000), Yleistetty parametrien suhteen lineaarinen mallirakenne ja signaalinkäsittely, Licentiate’s thesis, Helsinki University of Technology, Espoo, Finland. In Finnish.

Paliwal, K. K. & Kleijn, W. B. (1995), *Speech Synthesis and Coding*, Elsevier Science Publ., Amsterdam, the Netherlands, chapter Quantization of LPC parameters, pp. 433–466. In (Kleijn & Paliwal 1995*c*).

Papoulis, A. (1985), ‘Predictable processes and Wold’s decomposition: A review’, *IEEE Trans. Acoust., Speech, and Signal Proc.* **ASSP-33**(4), 933–938.

Pedersen, J. A., Rubak, P. & Tyril, M. (1999), Digital filters for low frequency equalization, *in* ‘AES 106th Convention preprint 4897’, Munich, Germany.

Priestley, M. B. (1981), *Spectral analysis and time series*, Vol. 2 of *Probability and mathematical statistics*, Academic Press Inc., London.

Purat, M. & Noll, P. (1996), Audio coding with a dynamic wavelet packet decomposition based on frequency-varying modulated lapped transforms, *in* ‘Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.’, Vol. 2, IEEE, Atlanta, USA, pp. 1021–1024.

Purnhagen, H., Edler, B. & Ferekidis, C. (1998), Object-based analysis/synthesis audio codec for very low bit rates, *in* ‘Proceedings of the 104$^{\text{nd}}$ Convention of the Audio Engineering Society, Preprint 4747’.

Rabiner, L. & Rader, C. M., eds (1972), *Digital Signal Processing*, Selected Reprint Series, IEEE Press, New York.

Ramprashad, S. A. (1999), A multimode transform predictive coder (MTPC) for speech and audio, *in* ‘Proc. IEEE Workshop on Speech Coding’, IEEE, Porvoo, Finland.

Robinson, E. A. (1982), ‘A historical perspective of spectrum estimation’, *Proc. IEEE* **70**(9), 885–907.

Roy, G. & Kabal, P. (1991), Wideband CELP speech coding at 16 kbits/s, *in* ‘Proc. of IEEE Int. Conf. on Acouctics, Speech, and Signal Processing’, Vol. 1, IEEE, Toronto, Canada, pp. 17–20.

Sarroukh, B. E. & den Brinker, A. C. (1998), Non-uniformly downsampled filter banks, *in* ‘Signal Processing IX: Theories and Applications’, Vol. I, EURASIP, Rhodes, Greece, pp. 265–268.

Savioja, L. (1999), Modeling techniques for virtual acoustics, PhD thesis, Helsinki University of Technology, Telecommunications Software and Multimedia Laboratory, Espoo, Finland. Report TML-A3.

Schafer, R. W. & Markel, J. D., eds (1979), *Speech Analysis*, Selected Reprints Series, IEEE Press, New York.

Scharf, B. (1970), Critical bands, *in* J. V. Tobias, ed., ‘Foundations of Modern Auditory Theory’, Academic Press, pp. 159–202.

Scheirer, E. D. (1999), Generalized audio coding with MPEG-4 Structured Audio, *in* ‘Proc. AES 17th Int. Conf.: High-Quality Audio Coding’, AES, Florence, Italy, pp. 61–75.

Schroeder, M. R. (1982), ‘Linear prediction, extremal entropy and prior information in speech signal analysis and synthesis’, *Speech Communication* **1**(1), 9–20.

Schroeder, M. R. & Atal, B. A. (1985), Code-excited linear prediction (celp): High quality speech at very low bit rates, *in* ‘Proc. of IEEE Int. Conf. on Acouctics, Speech, and Signal Processing’, pp. 937–940.

Schroeder, M. R., Atal, B. S. & Hall, J. L. (1979), ‘Optimizing digital speech coders by exploiting masking properties of the human ear’, *J. Acoust. Soc. Am.* **66**(6), 1647–1652.

Schüssler, W. (1970), ‘Variable digital filters’, *Arch. Elek. Übertragung* **24**, 524–525.

Singhal, S. (1990), High quality audio coding using multipulse LPC, *in* ‘Proc. Int. Conf. Acoust., Speech, and Signal Proc.’, Vol. I, IEEE, Albuquerque, USA, pp. 1101–1104.

Slepian, D. (1978), ‘Prolate spheroidal wave functions, Fourier analysis and uncertainty-V: The discrete case’, *Bell Syst. Tech. J.* **57**(5), 1371–1430.

Slutsky, E. (1927), ‘The summation of random causes as the source of cyclic processes’, *Problems of economic conditions*. Cited from (Wold, 1938).

Smith, J. O. & Abel, J. S. (1995), The Bark bilinear transform, *in* ‘Proc. IEEE ASSP Workshop’, New Paltz.

Smith, J. O. & Abel, J. S. (1999), ‘Bark and ERB bilinear transform’, *IEEE Trans. Speech and Audio Processing* **7**(6), 697–708.

Smith, J. O. & Serra, X. (1987), PARSHL: An analysis/synthesis program for nonharmonic sounds based on sinusoidal representation, *in* ‘Proc. Int. Computer Music Conf.’, pp. 290–297. Cited from (Verma 1999).

Sorenson, H. W. (1980), *Parameter Estimation: Principles and Problems*, Vol. 9 of *Control and Systems Theory*, Marcel Dekker, inc., New York.

Soulodre, G. A. & Lavoie, M. C. (1999), Subjective evaluation of large and small impairments in audio codecs, *in* ‘Proc. AES 17th Int. Conf.: High-Quality Audio Coding’, AES, Florence, Italy, pp. 329–336.

Soulodre, G. A., Grusec, T., Lavoie, M. & Thibault, L. (1998), ‘Subjective evaluation of state-of-the-art two-channel audio codecs’, *J. Audio Eng. Soc.* **46**(3), 164–177.

Stoica, P. & Moses, R. (1997), *Introduction to Spectral Analysis*, Prentice-Hall, New Jersey, USA.

Stoll, G. (1996), ISO-MPEG-2 audio: A generic standard for the coding of two-channel and multichannel sound, *in* N. Gilchrist & C. Grewin, eds, ‘Collected Papers on Digital Audio Bit-Rate Reduction’, Audio Engineering Society Inc., pp. 43–53.

Strube, H. W. (1980), ‘Linear prediction on a warped frequency scale’, *J. Acoust. Soc. Am.* **68**(4), 1071–1076.

Subba Rao, T. (1970), ‘The fitting of non-stationary signals’, *J. R. Statis. Soc.* **B32**, 312–322.

Szczupak, J. & Mitra, S. K. (1975), ‘Detection, location, and removal of delay-free loops in digital filter configurations’, *IEEE Trans. Acoustics, Speech and Signal Proc.* **23**(6), 558–562.

Tokuda, K., Kobayashi, T. & Imai, S. (1995), ‘Adaptive cepstral analysis of speech’, *IEEE Trans. Speech and Audio Processing* **3**(6), 481–489.

Tokuda, K., Kobayashi, T., Imai, S. & Chiba, T. (1993), ‘Spectral estimation of speech by mel-generalized cepstral analysis’, *Electr. and Comm. in Japan, Part 3* **76**(2), 30–43.

Tolonen, T. (2000), Object-based sound source modeling, PhD thesis, Helsinki University of Technology, Espoo, Finland. Lab. of Acoust. Audio Signal Proc., Report 55.

Toy, M. & Chirlian, P. M. (1984), ‘Low multiplier coefficient sensitivity block digital filters’, *IEEE Trans. Circuits and Systems* **CAS-31**(12), 993–1001.

Tsutsui, K., Suzuki, H., Shimoyoshi, O., Sonohara, M., Akagiri, K. & Heddle, R. M. (1996), ATRAC: Adaptive transform acoustic coding for MiniDisc, *in* N. Gilchrist & C. Grewin, eds, ‘Collected Papers on Digital Audio Bit-Rate Reduction’, Audio Engineering Society Inc., pp. 95–101.

Vaalgamaa, M., Härmä, A. & Laine, U. K. (1999), Audio coding with auditory time-frequency noise shaping and irrelevancy reducing vector quantization, *in* ‘Proc. AES 17th Int. Conference: High-Quality Audio Coding’, Florence, Italy, pp. 182–188.

Vaalgamaa, M., Härmä, A. & Laine, U. K. (2000), Subjective evaluation of LSF quantization in conventional and warped LP based audio coding, *in* ‘Signal Processing X: Theories and Applications’, EURASIP, Tampere, Finland, pp. 2065–2068.

Varho, S. (2001), New linear predictive methods for digital speech processing, PhD thesis, Helsinki University of Technology.

Varho, S. & Alku, P. (1999), A new predictive method for all-pole modeling of speech spectra with a compressed set of parameters, *in* ‘Proc. IEEE int. Symp. Circuits and Systems’, Orlando, Florida, pp. 126–129.

Verma, T. (1999), A perceptually based audio signal model with applications to scalable audio compression, PhD thesis, Stanford University, Stanford, USA.

Vernon, S. (1999), Dolby Digital: Audio coding for digital television and storage applications, *in* ‘Proc. AES 17th Int. Conf.: High-Quality Audio Coding’, AES, Florence, Italy, pp. 40–57.

Viswanathan, R. & Makhoul, J. (1975), ‘Quantization properties of transmission parameters in linear predictive systems’, *IEEE Trans. Acoust., Speech, Signal Processing* **ASSP-23**, 309–321. Reprinted in (Schafer & Markel 1979).

Wahlberg, B. (1991), ‘System identification using Laguerre models’, *IEEE Trans. Automatic Control* **36**(5), 551–562.

Wahlberg, B. (1994), ‘System identification using Kautz models’, *IEEE Trans. Automatic Control* **39**(6), 1276–1282.

Walker, G. (1931), ‘On periodicity in series of related terms’, *Proc. R. Soc. ser. A* **313**, 518–532.

Whittle, P. (1954), ‘Some recent contributions to the theory of stationary processes’, Appendix 2 in (Wold 1954).

Widrow, B. & Hoff Jr., M. E. (1960), Adaptive switching circuits, *in* ‘IRE WESCON Conv. Rec.’, pp. 96–104. Cited from (Haykin 1996).

Wiener, N. (1930), ‘Generalized harmonic analysis’, *Acta Math.* **55**, 117–258. Cited from (Robinson 1982).

Wiener, N. (1949), *Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications*, Technology Press and John Wiley & Sons, Inc., New York.

Wold, H. (1954), *A Study in the Analysis of Stationary Time Series*, 2 edn, Almquist and Wiksell, Stockholm, Sweden. first edition in 1938.

Yatrou, P. & Mermelstein, P. (1988), ‘Ensuring predictor tracking in ADPCM speech coders under noisy transmission conditions’, *IEEE J. on Sel. Areas in Comm.* **6**(2), 249–261.

Yule, G. U. (1927), ‘On a method of investigating periodicities in disturbed series with special references to Wolfer’s sunspot numbers’, *Philos. Trans. R. Soc. London, ser. A* **226**, 267–298.

Yule, G. U. & Kendall, M. G. (1958), *An introduction to the theory of statistics*, 14 edn, Charles Griffin & Co. LTD, London.

Zelinski, R. & Noll, P. (1977), ‘Adaptive transform coding of speech signals’, *IEEE Trans. Acoust. Speech, and Signal Processing* **ASSP-25**, 299–309.

Zhao, K., Ling, F., Lev-Ari, H. & Proakis, J. G. (1994), ‘Sliding window order-recursive least-squares algorithms’, *IEEE Trans. Signal Proc.* **42**(8), 1961–1972.

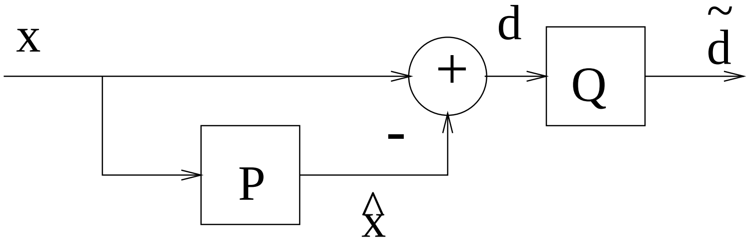
Zwicker, E. & Fastl, H. (1990), *Psychoacoustics: facts and models*, Springer-Verlag.

Zyda, M. J. (1992), ‘The software required for the computer generation of virtual environments’, *J. Acoust. Soc. Am.* **92**(4), 2457–2458.

Errata

Publication [P1]:

- The left part of Fig. 4 (WLP ENCODER) is wrong. A closed-loop encoder is shown in the Figure while an open loop structure was used. The correct structure is shown below:



5. Publications