Helsinki University of Technology Laboratory of Acoustics and Audio Signal Processing
Espoo 2001 Report 58

NEW LINEAR PREDICTIVE METHODS FOR DIGITAL SPEECH PROCESSING

Susanna Varho

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 58

NEW LINEAR PREDICTIVE METHODS FOR DIGITAL SPEECH PROCESSING

Susanna Varho

Dissertation for the degree of Doctor of Philosophy 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 20th of April, 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 951-22-5409-3
ISSN 1456-6303

Otamedia Oy
Espoo, Finland 2001

Abstract

Speech processing is needed whenever speech is to be compressed, synthesised or recognised by the means of electrical equipment. Different types of phones, multimedia equipment and interfaces to various electronic devices, all require digital speech processing. As an example, a GSM phone applies speech processing in its RPE-LTP encoder/decoder (ETSI, 1997). In this coder, 20 ms of speech is first analysed in the short-term prediction (STP) part, and second in the long-term prediction (LTP) part. Finally, speech compression is achieved in the RPE encoding part, where only 1/3 of the encoded samples are selected to be transmitted

This thesis presents modifications for one of the most widely applied techniques in digital speech processing, namely linear prediction (LP). During recent decades linear prediction has played an important role in telecommunications and other areas related to speech compression and recognition. In linear prediction sample *s(n)* is predicted from its p previous samples by forming a linear combination of the p previous samples and by minimising the prediction error. This procedure in the time domain corresponds to modelling the spectral envelope of the speech spectrum in the frequency domain. The accuracy of the spectral envelope to the speech spectrum is strongly dependent on the order of the resulting all-pole filter. This, in turn, is usually related to the number of parameters required to define the model, and hence to be transmitted.

Our study presents new predictive methods, which are modified from conventional linear prediction by taking the previous samples for linear combination differently. This algorithmic development aims at new all-pole techniques, which could present speech spectra with fewer parameters.

Acknowledgements

PhD-students are usually very happy already at this stage of the curriculum. However, I feel exceptionally happy and privileged to have had such wonderful people behind this project. Firstly, I would like to express my enormous gratitude to my supervisor professor Paavo Alku. He has had a clear view, great ideas and an infinite amount of time and patience to guide me through the study of predictive algorithms which has resulted in this PhD-thesis. In addition to his appreciated professional and pedagogical skills, I also highly value his politeness and friendliness in our co-operation.

I would like to thank the personnel of the Electronics and Information Technology Laboratory at the University of Turku together with the personnel of the Acoustics and Audio Signal Processing Laboratory at the Helsinki University of Technology for the supportive and encouraging working atmosphere. I have always felt welcomed in both research groups.

This research has been mainly funded by the Finnish Academy, through the Graduate School of Electronics, Telecommunication and Automation (GETA). I thank the Academy for giving me this opportunity to research, and especially the personnel of GETA for their flexibility and care of the students scattered over Finland. I would also like to thank the Electronics and Information Technology Laboratory at the University of Turku, the Acoustics and Audio Signal Processing Laboratory at the

Helsinki University of Technology, the Technology Development Foundation of the Finnish Ministry of Trade and Technology, the Sonera Research and Education Foundation and Elisa Communications for their financial support of my research.

For the encouraging comments, fruitful questioning and very valuable suggestions I would like to thank both pre-examiners of this thesis, Dr. Marko Juntunen and professor Jyrki Joutsensalo.

Finally, I would like to acknowledge my thankfulness to all the people who have been in this work in spirit. My family has enabled and encouraged me to pursue the curriculum, and most of all I thank my mother for making the impossible possible. Due to her efforts, example and insight added to her strong walking beside me in life, I would like to dedicate this book to her. We call this book her first grandchild.

Table of Contents

List of Publications

This thesis consists of an introduction and the following publications that are referred to by [P1], [P2], … [P9] in the text:

[P1] Varho, S. and Alku, P. 1997. A Linear Predictive Method Using Extrapolated Samples for Modelling of Voiced Speech, *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, Session IV, pp. 13-16.

[P2] Varho, S. and Alku, P. 1998a. Separated Linear Prediction - A new all-pole modelling technique for speech analysis, *Speech Communication*, **24**: 111-121.

[P3] Varho, S. and Alku, P. 1998b. Regressive Linear Prediction with Triplets - An Effective All-Pole Modelling Technique for Speech Processing, *Proceedings of IEEE International Symposium on Circuits and Systems*, Monterey, CA, Vol. IV, pp. 194-197.

[P4] Varho, S. and Alku, P. 1998c. Spectral Estimation of Voiced Speech with Regressive Linear Prediction, *Proceedings of IX European Signal Processing Conference*, Rhodes, Greece, Vol. II, pp. 1189-1192.

[P5] Alku, P. and Varho, S. 1998d. A New Linear Predictive Method for Compression of Speech Signals, *Proceedings of the 5th International Conference on Spoken Language Processing*, Sydney, Australia, Vol. VI, pp. 2563-2566.

[P6] Varho, S. and Alku, P. 1999. A New Predictive Method for All-Pole Modelling of Speech Spectra with a Compressed Set of Parameters, *Proceedings of IEEE International Symposium on Circuits and Systems*, Orlando, FL, Vol. III, pp. 126-129.

[P7] Varho, S. and Alku, P. 2000a. A Linear Predictive Method Highly Compressed Presentation of Speech Spectra, to be published in *Proceedings of IEEE International Symposium on Circuits and Systems*, Geneva, Switzerland, Vol. V, pp. 57-60.

[P8] Varho, S. and Alku, P. 2000b. Linear Prediction of Speech by Sample Grouping, *Proceedings of IEEE Nordic Signal Processing Symposium*, Kolmården, Sweden, pp. 113-116.

[P9] Varho, S. and Alku, P. 2000d. Separated Linear Prediction - Improved Spectral Modelling by Sample Grouping, *to be published in Proceedings of IEEE International Symposium on Intelligent Signal Processing and Communication Systems*, Honolulu, HI, pp. 731-735.

List of Abbreviations

AR	Autoregressive
ARMA	Autoregressive Moving Average
CELP	Code Excited Linear Prediction
DAP	Discrete All-Pole Modeling
DSP	Digital Signal Processing
EEG	Electroencephalogram
ETSI	European Telecommunication Standards Institute
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
F_0	fundamental frequency
GETA	Graduate School for Electronics, Telecommunication and Automation
GSM	Global System for Mobile communications
IIR	Infinite Impulse Response
IRLS	Iterative Reweighted Least Squares
IS	Itakura-Saito spectral flatness
LAR	log-area-ratio
LP	Linear Prediction
LPC	Linear Predictive Coding
LPES	Linear Prediction with Extrapolated Samples
LPLE	Linear Prediction with Linear Extrapolation
LPSG	Linear Prediction with Sample Grouping
LTP	Long-Term Prediction
L_1	least absolute error criterion
L_2	least squares error criterion
MA	Moving Average

NLP	Nonlinear Prediction
RBLP	Robust Linear Prediction
RLP	Regressive Linear Prediction
RLPT	Regressive Linear Prediction with Triplets
RPE	Regular Pulse Excitation
SER	Signal-to-Error Rate
SLP	Separated Linear Prediction
SSLP	Sample-Selective Linear Prediction
STP	Short-Term Prediction
TSP	Two-Sided Prediction
VQ	Vector Quantisation
WLP	Weighted Linear Prediction

List of Symbols

a_k	prediction coefficients of the past samples, i.e., outputs of the ARMA model
$A(z)$	inverse filter
b_l	prediction coefficients of the inputs of the ARMA model
b_i	backward prediction error
$e(n)$	error signal, residual
E	energy of the residual
E_{IS}	Itakura-Saito spectral flatness measure
$E(z)$	driving function, z-transform of $e(n)$
f_i	forward prediction error
G	prediction gain
$G(z)$	glottal shaping model
$H(k)$	spectrum of the all-pole model
$H(z)$	system function
k_i	reflection coefficients
M	number of steps on the unit circle
n	integer variable denoting time instant
N	interval of the predictive analysis in time
$L(z)$	lip radiation model
p	number of parameters defining the denominator of the predictive filter
$P(\omega_m)$	power spectrum of $s(n)$
$\hat{P}(\omega_m)$	power spectrum $\tilde{s}(n)$
q	number of parameters defining the numerator of the predictive filter
$R(i)$	autocorrelation coefficients
$\hat{R}(i)$	autocorrelation coefficients of the predictive filter

$R(z)$	lip radiation model
$s(n)$	discrete-time signal to be modelled
$\tilde{s}(n)$	prediction of sample $s(n)$
$s(t)$	continuous-time signal
$S(k)$	speech spectrum
$S(z)$	z-transform of $s(n)$
T	sampling interval
$u(n)$	input of the hypothetical system
$U(z)$	z-transform of $u(n)$
v_k	speech segment vector
V_p	normalised prediction error
$V(z)$	all-pole vocal tract model
$w(n)$	window function
W	weighting matrix
δ	threshold
φ_{ki}	covariance coefficients
ω_m	integer frequency variable

1. Introduction

Speech processing is a diversified area of research, primarily due to the fact that speech processing will always be needed in such important areas as telecommunication technologies. For example, speech processing in real-time and transparent speech transmission sets many challenges for researchers. The requirements for the quality of processed speech are increasing constantly. Since speech has many aspects, distinguishing between different utterances is not enough, all the other information related to e.g., identification of the speaker, the speaker's disposition or other emphases included in the speech signal should also be delivered to the receiver. In phonetics, the characteristics of speech, including intonation, enunciation and length of the phoneme, i.e., quantitety, are all called the prosody of the speech. Depending on the application and the required quality, all the perceptually important characteristics of speech must be maintained during processing. Future wireless or mobile communication applications, for example, will no longer tolerate speech quality which is inferior to the toll quality (Rabiner and Schafer, 1978; Kondoz, 1995).

In real-time speech processing the computation and delays should be minimised. However, in speech coding, for example, due to the transmission of speech, other resources of the system also become more limited. In addition to the computational complexity and delays, the bandwidth sets its limitations to the amount of data transmitted, i.e. to the number of parameters used and to the accuracy of quantisation applied. This means constant new demands for signal processing, solutions to be found for, among other things, telecommunication and multimedia applications.

1.1 Modelling of speech production

Before going into speech processing, some physical background for the speech signal itself can be presented. Sounds are usually categorised into three different groups: voiced sounds, unvoiced sounds and stop-consonants (Rabiner and Schafer, 1978). Vowels (e.g., /a/, /e/, /i/) and nasals (e.g., /n/, /m/) are voiced sounds, fricatives (e.g., /s/, /h/, /f/) are examples of unvoiced sounds, and plosives (e.g., /k/, /p/, /t/) belong to the stop-consonants group.

Physical speech production originates in the lungs, where the airflow begins. The actual sound is formed while air flows through the larynx and vocal tract (Parsons, 1986). The larynx consists of the cricoid cartilage, vocal folds and arytenoid cartilage. The vocal tract can be divided into three areas: oral pharynx, nasal cavity and mouth. The tongue, velum, lower jaw and lips have the most effect on the form of the vocal tract and therefore to the distinction of sounds. Speech production organs are presented in Fig. 1.

Depending on the activity status of the vocal folds, the sounds can be roughly divided into two extreme cases (Fant, 1960). First, the vocal folds can open and close periodically, generating in this way a train of pulses (glottal pulse). This gives the sound its voiced nature, i.e. periodicity in time and harmonic structure in frequency.

Second, the vocal folds can just be open, with the airflow forming turbulence between the folds and the sound very noise-like. Unvoiced sounds are formed this way. Stop-consonants are generated at the instant the vocal folds close.

Speech modelling can be divided into two different concepts: waveform coding and source coding (Deller et al., 1993). Initially, people have tried to imitate all the phenomena and their signals as they are; in speech processing this corresponds to waveform coding. In waveform coding there is an attempt to maintain the original waveform and the coding is based on quantisation and redundancy within the waveform. On the other hand people have a tendency to break the target of study into smaller components, and to analyse and model it piecewise. From these pieces we collect information on the subject of interest. This leads to the concept of source coding, which in turn aims at modelling speech with different parameters.

The source-filter theory or parametric speech production model is commonly presented as follows (Fant, 1960):

$$S(z) = E(z)G(z)V(z)L(z) \tag{1.1}$$

where *E(z)* is the driving function for the glottal shaping model *G(z)*, *L(z)* is the lip radiation model and *V(z)* is the vocal tract model. Typically in speech coding *G(z)*, *V(z)* and *L(z)* are all combined into the vocal tract model. The *E(z)* in the time domain *e(n)* is called the source excitation and it can be either a train of pulses or random noise. The excitation of the voiced sounds is usually modelled by the train of pulses, whereas unvoiced sounds use random noise excitation. However, other concepts for glottal excitation have also been presented, e.g., sinusoidal coding (Kleijn and Paliwal, 1995) models the excitation by the sinusoidal components of particular amplitudes, frequencies and phases.

Speech can be considered as a locally stationary process. However, parts of speech, i.e., voiced sounds can be considered as deterministic signals (Makhoul, 1975). Speech signal within a particular period of time follows a pattern, where

sample *s(n)* depends on its neighbouring samples, i.e., the samples correlate with each other. Therefore sample *s(n)* could be predicted from its previous samples, or *s(n)* could be presented by samples *s(n-k)*, $1 \leq k \leq p$. This feature of locally stationary signals has been exploited in linear prediction (LP), where sample *s(n)* is presented as a linear combination of its preceding samples. The prediction is optimised by minimising the prediction error (Atal and Hanauer, 1971), i.e. the prediction coefficients a_k are determined (see Chapter 2.3). Depending on the parameters of prediction, different characteristics of a speech segment can be modelled with LP. If the prediction takes into account a large number of samples the resulting model is very accurate and the error signal, the residual, becomes very low-energetic white noise.

In Fig. 2 the block diagram of Rabiner and Schafer's discrete-time model for speech production after is presented (Rabiner and Schafer, 1978). The figure can also be seen as an illustration of the applicability of linear prediction for speech modelling. This model is also called terminal-analog model, which means that the signals and systems involved are only superficially analogous (Oppenheim and Schafer, 1989). The vocal tract model *H(z)* and lip radiation model *R(z)* are excited by a discrete glottal excitation signal *u(n)*. For voiced speech the source excitation is an impulse train generator driving a glottal shaping filter *G(z)* and using local pitch period estimation. For unvoiced speech the source excitation comes from a random noise generator. However, this model omits the cases with more than one source of excitation (e.g. voiced fricatives) (Rabiner and Schafer, 1978). Hence, it can be concluded that LP does not match exactly to the speech production model presented by Fant (Fant, 1960).

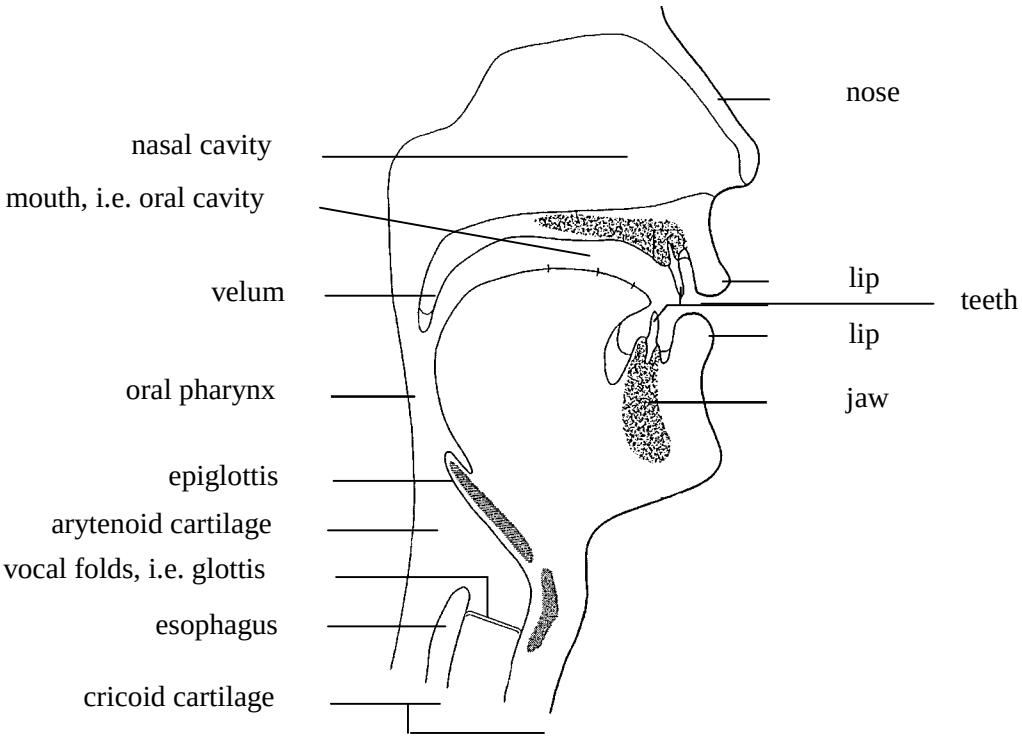


Figure 1. Speech production organs. Modified slightly according to Parsons (1986).

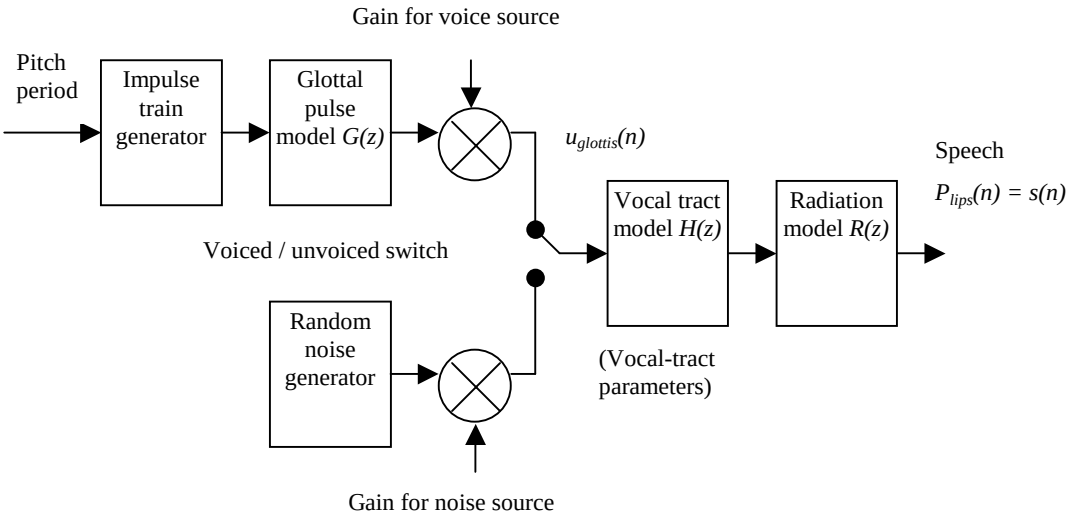


Figure 2. A general discrete-time model for speech production. After Rabiner and Schafer (1978).

2. Linear Prediction

2.1 Method for Time-Series Analysis

Linear prediction is a method which predicts the nth sample of the signal, *s(n)*, by forming a linear combination of p previous samples of *s(n)*. The linear combination is usually optimised by minimising the square of the prediction error. Linear prediction (LP) or linear predictive coding (LPC) was originally developed by mathematicians and physicists who studied time series analysis in the early decades of the 20th century. The first known application of linear prediction was in the analysis of sunspots by Yule (Yule, 1927). Since discrete-time speech processing has become increasingly important, also linear prediction has been implemented in various applications of DSP. Numerous studies of LP have resulted in effective computation algorithms and hence increased usability of LP.

Linear prediction forms a model of the signal in the time domain. This is especially useful if the signal is to be compressed. It may be more efficient, robust or otherwise reasonable to transmit the information of the signal model together with the error between the model and the actual signal instead of the signal itself. Presumably the model can be parameterised in a robust way and the error signal can be reduced to an approximation of, for example, white noise. This implies that the signal entropy of

the transmitted (error) signal can be decreased and therefore fewer quantisation levels are required.

Despite the formulation of linear prediction in the time domain, LP has interpretations in the frequency domain as well as in terms of the autocorrelation functions (Marple, 1987; Kay, 1988). Linear prediction has been studied in different areas in science, usually under different titles and terms, yet all aiming at describing either the signal or its spectrum with a relatively simple model, and therefore applying information about the signal source for the purpose in question.

The assumption of linearity is rarely true with real-life signals. However, a good linear model as a result of its simplicity and good behaviour, gives many benefits over the more complex nonlinear model, which may yield slightly better matching to the original problem, but cause much more difficulty in computation, memory usage or other processing resources. Besides, implementation of any model in real-life will assure nonlinearity to some extent even in theoretically linear systems. In some cases this nonlinearity can be exploited, whereas sometimes its effects should be minimised or masked. Linear prediction is basically a simple procedure, yet offering flexibility to develop and adapt the method according to the requirements of the specific field of interest. Following are some examples of these different scientific areas where LP has been applied.

After the sunspot analysis as the first application of linear prediction neuroscientists among others have used spectral estimation in studying electroencephalograms (EEG-signals) of the brain (Fenwick et al., 1971). Electric responses are measured from the brain and the spectrum of this EEG-signal is studied by modelling its basic characteristics with linear prediction. In addition to this LP is widely applied in other biosignal analyses, e.g. in heart rate variability studies (Baselli et al. 1985).

The surface of the earth is studied with seismographs. An explosion causes vibrations in the earth and these seismic traces are then measured and studied as impulse responses of different layers in the earth. Linear prediction offers a concept for deconvolution in order to obtain the desired impulses from the measured data (Makhoul, 1975).

Computerised systems have become so complex that there is need for modelling these systems itself in order to better control and manage them. The computational load has to be monitored and optimised between computing units. Another application of linear prediction is its usage in host load prediction (Dinda and O'Hallaron, 1999). There the running time of a certain task on a host can be predicted by LP. Hence, smaller confidence intervals between different tasks can be achieved and the overall efficiency of the system can be increased.

Economists have exploited the developments of time-series analysis; linear prediction can be applied to modelling rates, stock values, demands and other indexes in the market (Cheung et al., 1996, Chou et al.,1996). Long and short-term predictors are valuable for market analysts (Weigend and Gershenfield, 1994).

Within the field of telecommunication linear prediction and its modifications have been used widely. In addition to source coding there are examples where LP has been exploited in the area of channel coding (Deneire and Slock, 1999). Earlier the signal source related to linear prediction has been voice, but image processing has also applied linear prediction in various applications (Öztürk and Abut, 1992; Höntsch and Karam, 1997; Marchand and Rhody, 1997). Especially in lossless image processing, compression is achieved by coding the difference signal between the original image and its prediction. This is due to the fact that the residual image can be coded with fewer bits than the original image. Nowadays linear prediction has been applied adaptively together with other signal processing methods in numerous lossless image coders (Golchin and Paliwal, 1997; Lee, 1999; Motta et al., 1999).

Finally, some of the most important implementations of linear prediction are in speech processing, and they will discussed in the next chapter.

2.2 Linear Prediction in Speech Processing

In general speech is not a deterministic signal, nor it is a stationary signal, however, it can be considered to be locally stationary. Consequently, linear prediction can be applied in many areas of speech processing (Makhoul, 1975). For example, linear prediction can be used in clinical practise for analysing voice disorders (Baken and Orlikoff, 1999). In the area of engineering LP is a basic method for speech analysis, for compression and coding (Atal and Hanauer, 1971) or for speech and speaker recognition (Choi et al., 2000; Bimbot et al., 2000). In speech analysis LP has been widely used in formant extraction in voice-source analysis (Alku, 1992).

2.3 Pole-Zero Model

Linear prediction of the nth sample *s(n)* is defined as the linear combination of p samples previous to sample *s(n)* (e.g., Atal and Hanauer, 1971). Hence the mathematical presentation for the prediction of *s(n)* denoted by $\tilde{s}(n)$ can be expressed as:

$$\tilde{s}(n) = \sum_{k=1}^{p} a_k s(n-k) \qquad (2.1)$$

where a_k, $1 \le k \le p$ denote for the prediction coefficients. This most commonly-used model, consisting only of previous samples or outputs *s(n)* of the system, can be generalised to include also the inputs *u(n)* of the system; therefore the presentation of signal *s(n)* according to this so-called Autoregressive Moving Average model (ARMA) is as follows (Makhoul, 1975):

$$s(n) = \sum_{k=1}^{p} a_k s(n-k) + G \sum_{l=0}^{q} b_l u(n-l) \tag{2.2}$$

where a_k, $1 \le k \le p$, b_l, $1 \le l \le q$, and G is the gain. The time-domain presentation in Eq. 2.2 can be expressed in frequency domain as follows:

$$H(z) = \frac{S(z)}{U(z)} = G \frac{1 + \sum_{l=1}^{q} b_l z^{-l}}{1 + \sum_{k=1}^{p} a_k z^{-k}} \tag{2.3}$$

where *S(z)* and *U(z)* are the z-transforms of *s(n)* and *u(n)*, respectively. *H(z)* is the system function of the system to be modelled, i.e. the pole-zero model. The roots of the numerator and the denominator yield the zeros and the poles of the system, respectively. This leads also to the special cases of this general model. If we approximate the numerator to be fixed (i.e., $b_l = 0$, $1 \le l \le q$), the model reduces to the all-pole model, or autoregressive (AR-) model (Marple, 1987; Kay, 1988):

$$H(z) = \frac{G}{1 + \sum_{k=1}^{p} a_k z^{-k}} \tag{2.4}$$

Correspondingly, the other special case can be derived, if the denominator is approximated as fixed (i.e., $a_k = 0$, $1 \le k \le p$), and the model becomes an all-zero model. This is also called the moving average, the MA-model. In terms of spectral modelling it can be epitomised that the all-pole model estimates basically the local maxima and the all-zero model matches primarily the local minima of the spectrum. In spectral modelling of speech, it is perceptually more important to match the most energetic parts of the spectrum, i.e., the local maxima or formants of speech, than spectral valleys, where there is only little energy. Therefore, speech processing usually takes advantage of all-pole models. All-pole modelling is a rather simple and

straightforward method which does not require a large amount of computation or memory, especially in comparison to the pole-zero model. Furthermore, the pole-zero model can not be solved in a closed form.

2.4 All-Pole Model

The error between the predicted sample and the actual sample, the residual is as follows:

$$e(n) = s(n) + \tilde{s}(n) = s(n) + \sum_{k=1}^{p} a_k s(n-k) \tag{2.5}$$

The signal at time instant n can then be reconstructed from the residual sample at time instant n and from the previous samples of *s(n)*:

$$s(n) = e(n) - \tilde{s}(n) = e(n) - \sum_{k=1}^{p} a_k s(n-k) \tag{2.6}$$

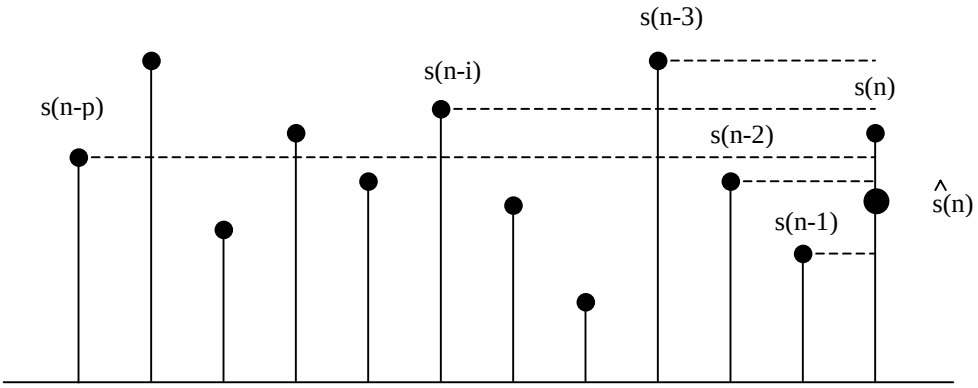


Figure 3. Linear prediction of sample *s(n)*

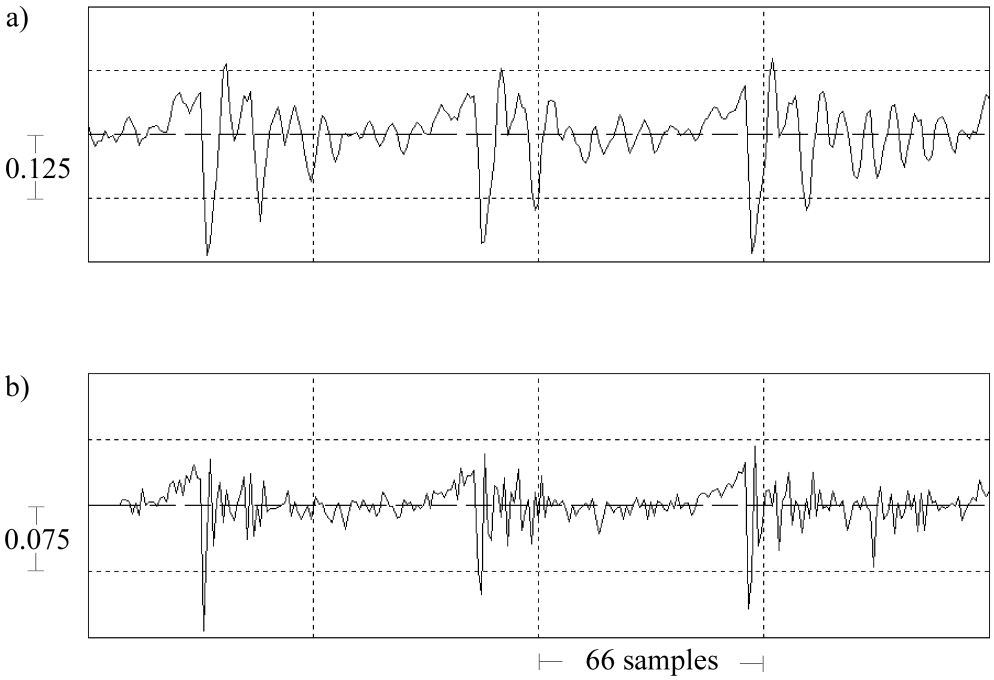


Figure 4. a) Speech signal (vowel /a/, male speaker), b) the residual given by linear prediction, p=8.

If the p-value is large enough, the all-pole model is able to model the spectral envelope of the speech so accurately that the residual becomes flat in spectrum. Mathematically this can be presented as follows:

$$E(z) = \left[1 + \sum_{k=1}^{p} a_k z^{-k}\right] S(z) = A(z)S(z) \tag{2.7}$$

$$\Rightarrow S(z) = \frac{E(z)}{A(z)} \approx \frac{G}{1 + \sum_{k=1}^{p} a_k z^{-k}} \tag{2.8}$$

where *E(z)* and *S(z)* are z-transforms of *e(n)* and *s(n)*, respectively. *A(z)* is called the inverse filter, while *S(z)* is the actual all-pole filter.

During recent decades linear prediction has been developed and modified, but the basic formulation presented here is called conventional linear prediction (Atal and Hanauer, 1971). The prediction in conventional LP is optimised by minimising the square of the prediction error:

$$E = \sum_n e^2(n) = \sum_n \left[s(n) + \sum_{k=1}^{p} a_k s(n-k) \right]^2 \tag{2.9}$$

Alternatively, the prediction error in the frequency domain is as follows:

$$E = \sum_n e^2(n) = \frac{1}{M} \sum_{m=0}^{M-1} \left| E(e^{j\omega_m}) \right|^2 = \frac{1}{M} \sum_{m=0}^{M-1} P(\omega_m) A(e^{j\omega_m}) A(e^{-j\omega_m}) \tag{2.10}$$

Minimisation is done by setting the partial derivatives of E with respect to a_i to zero, i.e.:

$$\frac{\partial E}{\partial a_i} = 0, \;\; 1 \le i \le p \,. \tag{2.11}$$

Optimisation of prediction according to the least squares criterion leads to the so-called normal equations:

$$\sum_{k=1}^{p} a_k \sum_n s(n-k)s(n-i) = -\sum_n s(n)s(n-i), \;\; 1 \le i \le p \tag{2.12}$$

If we assume the error to be minimised over an infinite duration of the signal, $-\infty < n < \infty$, then the solution of the normal equations can be obtained by the autocorrelation method. The autocorrelation function is defined as follows:

$$R_n(i) = \sum_{m=-\infty}^{\infty} s_n(m) s_n(m+i) \tag{2.13}$$

Applying the autocorrelation function results in the following normal equations:

$$\begin{bmatrix} R_n(0) & R_n(1) & R_n(p-1) \\ R_n(1) & R_n(0) & R_n(p-2) \\ \\ R_n(p-1) & R_n(p-2) & R_n(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \\ a_p \end{bmatrix} = -\begin{bmatrix} R_n(1) \\ R_n(2) \\ \\ R_n(p) \end{bmatrix} \tag{2.14}$$

These equations are also called the Yule-Walker equations (Yule, 1927; Walker, 1931). The autocorrelation method yields a matrix, the diagonal elements of which are equal; i.e. the resulting matrix is a Toeplitz matrix; additionally the matrix is symmetric. Hence, the solution for the matrix equation can be obtained quickly, for example by using the Levinson-Durbin recursion (Makhoul, 1975; Marple, 1987). The assumption of the signal to be infinite must be corrected, since in practice the signal is known only on a finite interval; this is taken care of by using a window function *w(n)*.

The covariance function can be applied if we assume the error to be minimised over a finite interval $0 \le n \le N-1$. Then optimisation of Eq. 2.12 reduces to:

$$\sum_{k=1}^{p} a_k \varphi_{ki} = -\varphi_{0i}, \quad 1 \le i \le p \tag{2.15}$$

where the covariance function is:

$$\varphi_{ki} = -\sum_{n=0}^{N-1} s(n-i)s(n-k) \tag{2.16}$$

Normal equations with the covariance method are as follows:

$$\begin{bmatrix} \varphi(1,1) & \varphi(1,2) & \varphi(1,p) \\ \varphi(2,1) & \varphi(2,2) & \varphi(2,p) \\ & & \\ \varphi(p,1) & \varphi(p,2) & \varphi(p,p) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \\ a_p \end{bmatrix} = - \begin{bmatrix} \varphi(1,0) \\ \varphi(2,0) \\ \\ \varphi(p,0) \end{bmatrix} \tag{2.17}$$

The covariance matrix is not a Toeplitz matrix, but it is also symmetric. However, most of the computational load in LP comes from solving for autocorrelation or covariance coefficients, not from matrix inversion. These coefficients require *pN* operations, while even the general matrix inversion methods require $p^3/6+O(p^2)$ operations, normally $N >> p$, therefore the issue of the matrix type is not critical. For comparison, a recursive Levinson-Durbin method, which can be applied to Toeplitz matrices, requires $p^2+O(p)$ operations (Makhoul, 1975). Another advantage for the

covariance method is that it can also be derived to be applicable for the linear prediction of non-deterministic signals. The covariance method does not suffer from the truncation of the signal, since it assumes the signal to be finite. One drawback of the covariance method is that it cannot be guaranteed to yield a stable all-pole filter, as is the case with the autocorrelation method for all nonzero signals ($E_i > 0$). Finally, it can be noted that the covariance method approaches the autocorrelation method as its block size approaches infinity.

2.5 Linear Prediction in Spectral Modelling

Linear prediction can be analysed either in the time or frequency domain, but also in terms of the autocorrelation functions. With respect to the relationship between LP and the autocorrelation function it is known that the p+1 first autocorrelation coefficients of the impulse response of the all-pole filter *H(z)* are exactly the same as the autocorrelation coefficients of the original signal (Makhoul, 1975; Rabiner and Schafer, 1978):

$$\hat{R}(i) = R(i), \quad 0 \le i \le p \,. \tag{2.18}$$

where $\hat{R}(i)$ and *R(i)* denote for the autocorrelation coefficients of the predictive filter and the autocorrelation function of the signal, respectively. This phenomenon can be seen in the frequency domain so that the resulting LP-model approaches the original spectrum as the p-value grows. As a limit $\hat{R}(i) = R(i)$ for all i as $p \to \infty$ and especially $\hat{P}(\omega_m) = P(\omega_m)$ as $p \to \infty$, and any spectrum can be approximated arbitrarily closely by an all-pole model, assuming that the autocorrelation estimate is errorless. On the other hand LP can be seen as a spectral smoothing method: As the p-value decreases the all-pole spectrum models fewer details of the original spectrum, and becomes more like the envelope of the spectrum or just represents the overall

tilting on the frequency scale. Matching of the all-pole model to the FFT spectrum of the speech signal with different values of p can be seen in Fig. 5.

In most applications there is always a compromise concerning the order of prediction, i.e. the number of samples to be taken into linear combination and hence also the number of unknowns in the normal equations or the number of parameters needed to characterise the all-pole model. A larger p, a more accurate spectral envelope, more computation, more delay and more bits for the transmission of parameters are needed. Moreover, as the order of prediction p increases the normalised error V_p (i.e., normalised autocorrelation coefficients result in normalised prediction error) decreases, and this implies greater risk for ill-conditioning (due to $p \approx N$). Therefore, p should be optimised according to the application. There are different criteria for optimising p. A very straightforward test is the following threshold test (Makhoul, 1975):

$$1 - \frac{V_{p+1}}{V_p} < \delta \tag{2.19}$$

where V_p and V_{p+1} denote normalised prediction errors with an order of prediction equal to p and p+1, respectively. The accepted threshold for the decrease in the prediction error is denoted by δ. Finally, the application itself describes best the optimisation criterion and thus a suitable order of prediction can be found. In coding applications the number of bits allocated to the quantisation of LP coefficients is always limited, hence typical values for p are 8, 10 or 12.

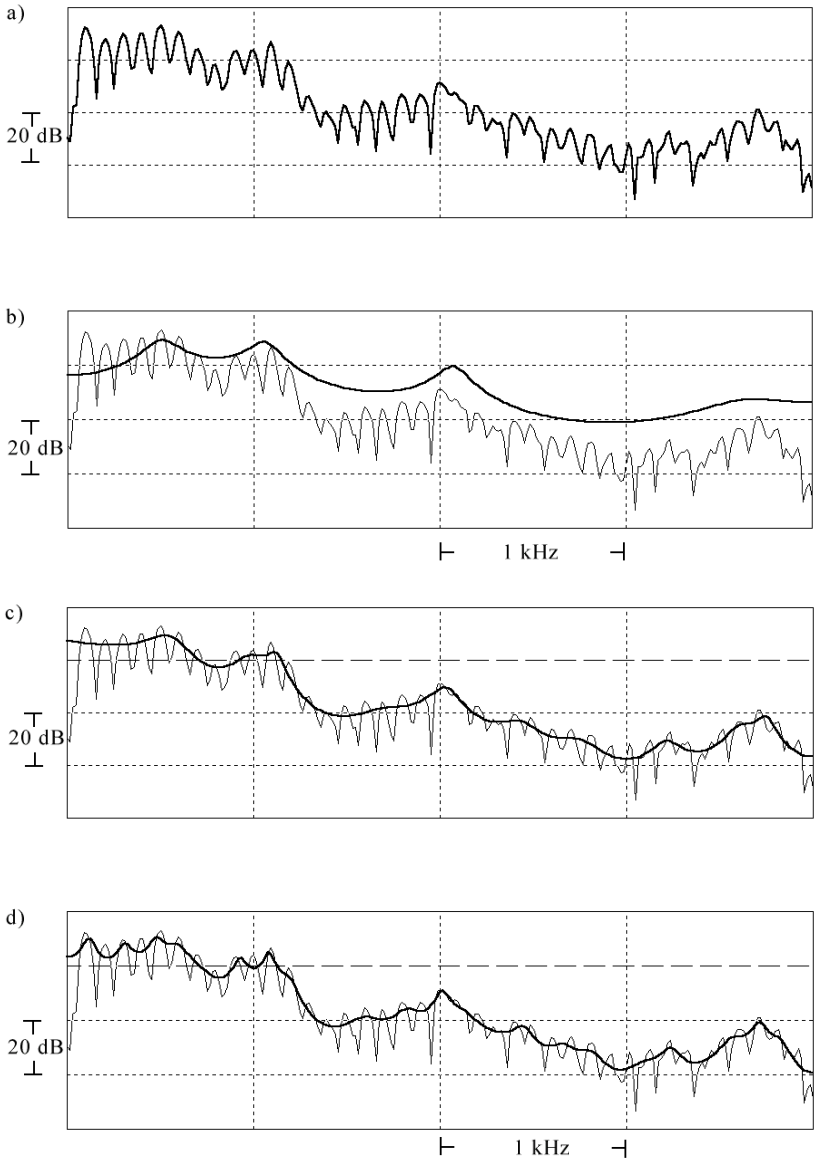


Figure 5. a) FFT-spectrum of vowel /a/, male speaker, b) LP: p=8, c) LP: p=28, d) LP: p=48.

All modelling techniques have their limitations, as does also linear prediction (Makhoul, 1975). Due to the mean square error criterion applied in linear prediction, LP-models tend to deteriorate: the lowest formants become biased towards high-energy peaks. This feature can be noticed when analysing the behaviour of LP in the modelling of voiced speech spectrum with harmonic structure. Low fundamental frequency F_0 is typical for male speakers, whereas women and children normally have high pitch, and thus much more sparse harmonic spectral structure in comparison to men. Linear prediction has difficulties in distinguishing formant peaks from harmonic peaks in particular when there are only a few harmonics in the analysis bandwidth.

One reason for the deterioration of linear prediction is that, according to the definition of LP, with a finite p there is always some error in spectral modelling, i.e. $E_p > 0$ for positive definite spectra. It can be understood that, for example, the impulse response of an all-pole filter can be otherwise predictable except for its initial nonzero value. The least squares error criterion is a reasonable choice, since it emphasises great errors while weighting only little small errors. From the time domain it is difficult to evaluate the applicability of the error criterion. However, spectral matching presents a model which more clearly fits the actual spectrum. The prediction error in the frequency domain is the ratio $E(\omega_m) = \frac{P(\omega_m)}{\hat{P}(\omega_m)}$; this leads to an even distribution of error on the frequency scale in general (Makhoul, 1975). This means that globally there is no distinction in spectral matching according to the energy on the average; the frequencies with high or low energy have on the average as good matching. However, locally the error on a small region is averaged, so that the arithmetic mean of $E(\omega_m)$ equals 1. Therefore the contribution of greater $E(\omega_m)$, i.e. $P(\omega_m) > \hat{P}(\omega_m)$ to the total error is larger than of smaller $E(\omega_m)$, i.e. resulting from $P(\omega_m) < \hat{P}(\omega_m)$. This feature of conventional linear prediction leads to the

spectral envelope concept, since the error criterion emphasises better matching in the frequencies where $P(\omega_m) > \hat{P}(\omega_m)$ (Makhoul, 1975).

Finally, spectral matching depends mostly on the spectrum that is to be modelled. Naturally, the all-pole model suits accurately for spectra where there are local resonances that can be modelled by poles. Local minima can also be modelled with poles, however requiring more than one pole for representing one zero. In the spectra of nasal sounds there are zeros; those zeros for which the bandwidths are wide can be more easily modelled with poles than narrow bandwidth zeros. The dynamics of the spectrum also affect linear prediction. It can be shown that large dynamics lead to small normalised prediction error V_p, which in turn leads to ill-conditioning of LP (Makhoul, 1975). There are various effective methods for avoiding ill-conditioning of LP, e.g., regularisation techniques, where presumed limitations are taken into account in the solving procedure (Gersho, 1996). In speech coding a common procedure to avoid ill-conditioning of linear prediction is the flattening of spectrum by filtering it with a first-order high-pass filter, the pre-emphasiser (Markel and Gray, 1976). This finite impulse response filter can be, for example, a first order predictor, but in speech coding applications a fixed FIR-filter with its zero typically close to the unit circle is widely used (ETSI, 1997). Pre-emphasis is cancelled with a corresponding first-order all-pole filter before reconstructing the signal at the receiver.

2.6 Computation of Parameters

Linear prediction removes redundancy from the signal. Therefore, it can be used effectively in the compression of data to be transmitted. In telecommunications applications both the linear prediction parameters and the residual need to be quantised and transmitted. There are theoretically many alternatives in presenting the model parameters (Makhoul, 1975). The impulse responses of the inverse filter *A(z)*,

a_k, can be quantised. Autocorrelation coefficients, spectral coefficients or cepstral coefficients as well as the poles themselves could be used for the model presentation in addition to the so-called reflection coefficients. However, in order to quantise the parameters they should be able to maintain the stability of the all-pole filter upon quantisation and they should have natural ordering. Natural ordering means that among the set of parameters the parameters have their own positions which are not interchangeable, i.e. a_1 can not be changed to a_2 or any other of the a_i, $i \neq 1$ without affecting the all-pole filter. Ordering does not exist for the poles.

Efficient presentation of linear prediction for quantisation is the use of reflection coefficients k_i. The reflection coefficients can be derived from the prediction coefficients a_i by transforming the prediction into the Lattice filter (Makhoul, 1975; Markel and Gray, 1976; Rabiner and Schafer, 1978). This recursion is as follows

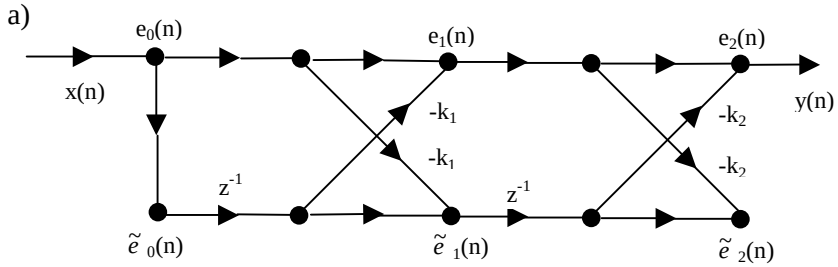
$$k_i = a_i^{(i)}, \quad a_m^{(i-1)} = \frac{a_m^{(i)} + k_i a_{i-m}^{(i)}}{1 - k_i^2}, \quad m = 1,2,...(i-1) \tag{2.20}$$

Another viewpoint for reflection coefficients is the following presentation of the normalised prediction error V_p yielded from Durbin's recursion:

$$V_p = \prod_{i=1}^{p} (1 - k_i^{\,2}), \quad 1 \leq i \leq p \tag{2.21}$$

Stability checking of reflection coefficients is straightforward. For a stable filter the roots of the denominator, i.e. all the poles, lie inside the unit circle. This corresponds to $|k_i| < 1$ for the reflection coefficients. The limited scale of k_i coefficients is beneficial in defining the quantisation scale: the robustness of coefficients increases and the stability upon quantisation can be maintained. Lattice structure is also feasible to implement due to its good numerical characteristics. The errors in the filter coefficients do not cumulate according to the increase of the Lattice filter order.

a)



b)

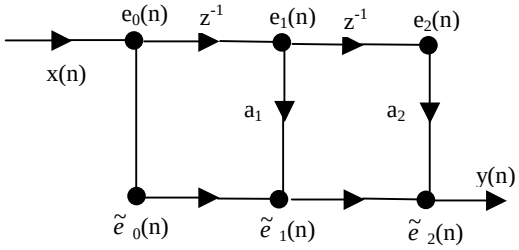


Figure 6. a) FIR Lattice filter of order 2

b) FIR direct form filter of order 2

3. Modifications of Linear Prediction in Speech Processing

Linear prediction is applied in various areas of speech processing. The basic formulation of linear prediction is simple enough; there are fast methods for the computation of LP with the autocorrelation method (the Levinson-Durbin recursion) and hence LP's applicability for speech processing is very efficient. But as mentioned in Chapter 2.5, conventional LP also has its drawbacks. One drawback of linear prediction is its tendency to deterioration towards pitch harmonics, especially for spectra with a sparse harmonic structure. Another drawback of LP originates with the restriction of the all-pole model compared to the pole-zero model: e.g. accurate modelling of nasal sounds would also require zeros in addition to the poles that the LP-synthesis filter yields (Rabiner and Schafer, 1978).

In order to overcome the drawbacks numerous developments and variations of linear prediction have been presented over the past decades (e.g. Atal and Schroeder, 1979; Strube, 1980; Marple, 1982; Hermansky, 1990; El-Jaroudi and Makhoul, 1991; Ma et al., 1993). In this chapter some of these different modifications of linear prediction are viewed.

3.1 Modification in Error Criterion of Optimisation

The difficulties of conventional linear prediction in modelling discrete spectra originate from the error criterion used in the optimisation (Makhoul, 1975; Denoël and Solvay, 1985; Alcázar-Fernández and Fraile-Peláez, 1988; El-Jaroudi and Makhoul, 1991). Minimising the error in Eq. 2.10 corresponds to fitting the autocorrelation of the LP synthesis filter to the aliased version of the autocorrelation of the original signal. This results in a deteriorated all-pole model of the original speech spectrum. In order to overcome this drawback of conventional linear prediction, El-Jaroudi and Makhoul presented an alternative method, Discrete All-Pole Modeling (DAP), for optimising prediction (El-Jaroudi and Makhoul, 1991). Their idea is to firstly approximate the speech spectrum by a line spectrum and then find a spectral envelope that best fits to this line spectrum. This procedure corresponds to the application of the Itakura-Saito spectral flatness measure as the error criterion. The Itakura-Saito spectral flatness measure can be defined as follows:

$$E_{IS} = \frac{1}{M}\sum_{m=1}^{M}\left[\frac{P(\omega_m)}{\hat{P}(\omega_m)} - \ln\frac{P(\omega_m)}{\hat{P}(\omega_m)} - 1\right] \tag{3.1}$$

This error criterion maximises the flatness of the residual spectrum, because spectral flatness is defined as the ratio between the geometric and arithmetic means of the spectral samples. For continuous spectra the Itakura-Saito spectral measure yields the same all-pole model as conventional linear prediction, therefore it can be noted that DAP reduces to conventional LP as the number of spectral points goes to infinity. Discrete All-Pole Modeling requires more computation as the solution for p+1 nonlinear equations must be found iteratively.

Denoël and Solvay presented a linear programming algorithm which optimises linear prediction according to the least absolute error (L_1) criterion (Denoël and Solvay, 1985). The least absolute error criterion is more robust than the least squares

error (L_2) criterion, because L_1 is less sensitive to high amplitude values than the least squares error criterion. The L_1 criterion suits best for problems where the distribution of prediction errors follows exponential distribution. Denoël and Solvay presented in their paper an algorithm that is based on Burg's method (Gray et al., 1977) for computing reflection coefficients, where the L_2 criterion in the method is replaced with the L_1 criterion. Computation of predictive parameters is much more complex (Denoël and Solvay, 1985) than in conventional linear prediction; however, the L_1 algorithm is presented as an efficient recursion.

Alcázar-Fernandez and Fraile-Peláez developed further the idea of applying the Burg algorithm together with some other error criterion than L_2 (Alcázar-Fernandez and Fraile-Peláez, 1988). Their proposal was to modify the original Burg's error, which is the sum of the squares of the forward and backward prediction errors (denoted by f_i and b_i, respectively) in the ith cell output by replacing the squares with the exponent p, $1 \le p \le 2$, as follows:

$$E_p = \sum_n \left| f_i(n) \right|^p + \left| b_i(n) \right|^p = \sum_n \left| f_{i-1}(n) + k_i b_{i-1}(n-1) \right|^p + \left| k_i f_{i-1}(n) + b_i(n-1) \right|^p \tag{3.2}$$

where the reflection coefficient k_i are solved iteratively as the forward prediction error. Hence, prediction is optimised according to L_p criterion $1 \le p \le 2$. An iterative reweighted least squares (IRLS) algorithm is used for solving reflection coefficients and p.

3.2 Modification from Perceptual Perspectives

In addition to modifying the error criterion of prediction, the perceptual aspects can be taken into account in speech modelling. Firstly, it is known that formant bandwidth increases with frequency. Secondly, the human ear has frequency resolution which is less sharp at higher frequencies, i.e., above 3.5 kHz (Strube, 1980), Therefore, in

order to improve the perceptual quality (Hermansky, 1990; Bourget et al., 1995) of speech model, the emphasis in speech modelling is set to lower, more critical frequencies, and the upper formants can be modelled with fewer poles, as they are naturally smoother and more sparse than the lower ones. In various modelling techniques that have been developed, frequency selectivity has been an advantageous feature (Reddy and Swamy, 1984; Hanson et al., 1994). One way to develop conventional linear prediction in the frequency domain is to replace the unit delays in the prediction with first-order all-pass filters (Strube, 1980; Laine, 1995). This yields a frequency warping property for linear prediction which can be exploited for controlling the modelling accuracy on the frequency scale. The frequency warping can be derived from the all-pass transformation (Strube, 1980).

3.3 Structural Modification

The structural modification of linear prediction can be done in the time domain. A straightforward method to modify the prediction is to take the samples for the linear combination differently. For example, in addition to the past samples of *s(n)*, also the future samples can be taken into prediction (Marple, 1982; Kay, 1983; Lee; 1989; Hsue and Yagle, 1993). David and Ramamurthi have also proposed a two-sided prediction model, TSP, which takes both p previous samples and p future samples into prediction (David and Ramamurthi, 1991). When applying the autocorrelation method the resulting all-pole filter is symmetric, requiring only half of its coefficients to be coded at the receiver. However, the TSP-filter can never be of minimum phase and hence the corresponding inverse filter is unstable. TSP was developed for frame-based prediction and it is derived by applying cyclic convolution.

3.4 Modification of Sample Selection

Miyoshi et al. have developed a method that selects within one frame samples for prediction that yield the relatively smallest prediction errors (Miyoshi et al., 1987). In the prediction of a voiced sample there may also be samples that are rated as unvoiced samples, hence having a different type of excitation (random noise) than the one to be predicted would have (train of impulses). Especially in voiced speech with a short pitch period this may cause problems. Miyoshi et al. have proposed a technique, Sample-Selective Linear Prediction (SSLP), to overcome this problem. SSLP has two stages: in the first stage SSLP computes the residuals for all the frame T_a, and in the second stage it selects the smaller range of samples T_w that yield smaller residuals than the set threshold θ, and finally applies these samples in T_w in the prediction. The proposed method models formants more accurately than conventional linear prediction, although the stability of the resulting filter cannot be guaranteed.

Sample-Selective Linear Prediction or its generalisation, Weighted Linear Prediction (WLP) (Kakusho et al. 1984) can also be seen as a special case of a robust M-estimate for the LP coefficients (Lee, 1988). Lee has studied the weighting of prediction residuals in this paper. Conventional linear prediction is based on the assumption that excitation for voiced sounds will have Gaussian distribution. Lee proposes an algorithm that relies on source excitation that has so called heavy-tailed non-Gaussian distribution. Mostly this mixture source (Lee, 1988) has Gaussian excitations, but a small portion of the excitations has Laplacian distribution with much larger variance. This leads to weighting more smaller residuals whereas larger residuals get less weight. The proposed method, Robust Linear Prediction (RBLP), is less sensitive to the length of the pitch period, the location of the excitation, or to windowing.

Already in the study presented by Steiglitz and Dickinson (Steiglitz and Dickinson, 1977) it was stated that the model of voiced speech can be improved if the

samples taken into prediction correspond to the open phase of the glottal cycle. This so-called undriven segment does not include the main excitation of the vocal tract. By modifying the prediction in such a way that voiced sounds are predicted from only a set of samples with the open glottis, the number of samples needed for prediction can be decreased. Steiglitz and Dickinson have proposed an algorithm, which modifies the segment of samples for prediction by first computing conventional LP from the set of samples. Then it finds the maximum point of the residual, k_{max}. Finally, the set of samples applied in the prediction is selected to be the samples preceding k_{max} after some agreed marginal of samples, e.g., from k_{max} +10 to k_{max} +266.

3.5 Nonlinear Prediction

Since speech is a non-linear real-life signal, a linear system can model it only arbitrarily closely. Just as numerous attempts to modify conventional linear prediction have been made, so also predictive methods which abandon linearity have been proposed to a great extent (Townshend, 1991; Wang et al., 1994; Birgmeier et al., 1997; Shimamura and Hayakawa, 1999).

As a generalisation of linear combination, the optimal least squares estimate of a random variable is its conditional expectation given by the observed variables. This leads to finding a multivariate probability density function of speech, which is not necessarily either analytically or computationally attractive (Wang et al., 1994). To simplify, it can also be seen that a nonlinear predictive method requires definition for its initial state, both definition for mapping the initial values to the predicted one, and finally training the system (Townshend, 1991). The accuracy of the nonlinear predictor depends on the structure of the predictor, the number of parameters to be applied, and the training data. The initial state of the predictor can be obtained from the past values of the signal. Mapping can be parametric or non-parametric, including e.g. layered neural networks, radial basis functions or linear prediction. Townshend

proposed in his paper (Townshend, 1991) a method with non-parametric mapping which applies local approximation. In local approximation mapping is broken into small neighbourhoods (i.e., *k* nearest neighbours in the table), where some parametric model is applied in each. Townshend's nonlinear predictor was trained on the residual of the conventional linear predictor. However, the implementation of this nonlinear predictor in a CELP coder (Kroon and Atal, 1987) would require an enormous amount of computation, thus the benefits in prediction accuracy remain insufficient.

Wang et al. presented (Wang, et al., 1994**)** another nonlinear predictive method that does not require any parametric model of the predictor. It maps straightforwardly from the quantised vector of past values to the predicted value by the means of table lookup. Wang's method, Nonlinear Prediction (NLP), applies standard vector quantisation (VQ) design algorithm for generating the codebook for the speech sequence vectors, v_k. According to their study NLP tends to also eliminate higher harmonics of the fundamental frequency and therefore flattens the residual spectrum more than conventional linear prediction. However, the codebook size for NLP should be rather large in order to exceed the prediction gain of LP. This would in turn inflict the need for larger memory and for increased computational complexity. Despite this, or due to this, the emphasis in speech analysis has recently been on nonlinear methods (Shimamura and Hayakawa, 1999).

4. Scope of the Thesis

In this thesis the modifications of linear prediction have been studied by emphasising the modelling of the spectral envelope of speech. Although the study can be regarded as basic research, possible applications in telecommunication, in particular in speech coding, have been contemplated. These applications set certain criteria and limitations that have been considered in developing algorithms. Publications P1 – P4, P6 and P8 included in this thesis present new predictive algorithms for the spectral estimation of speech; in publications P5, P7 and P9 quantisation characteristics of the proposed algorithms are studied.

4.1 Reformulation of Linear Prediction by Sample Grouping

In this thesis the emphasis has been on the development of algorithms that enable an accurate modelling of all-pole spectra with a compressed set of parameters. This means that with a given number (p) of parameters allocated to determine an all-pole filter, the new methods yield all-pole filters whose orders are larger than p. This, in turn, yields more accurate spectral modelling characteristics in comparison to conventional linear prediction. More accurate models of speech spectra are reflected especially in better matching of formants. Modelling one formant always requires a complex conjugate pair of poles for the all-pole filter. Therefore, the all-pole filter should have more than twice as many poles as there are formants to be modelled. As

the telephone bandwidth of speech contains four to five formants, the all-pole filter should have at least ten poles, and yet all the poles should occur as complex conjugate pairs in order to match five formants. The basic idea in our study has been to lengthen the all-pole filter by taking more past samples into prediction, but in order to hold to the given number of parameters, the past samples need to be combined in a specific way.

Another issue underlying the modification of LP has been the fact that the neighbouring samples have the largest correlation with the sample to be predicted *s(n)*, (Rabiner and Schafer, 1978). Surely, conventional linear prediction takes also this into account and optimises the a_k coefficients in a way that weights more the less-delayed samples. However, we have also studied the effect of additional emphasising of the closest samples to the predicted one.

Consequently, we have developed a group of linear predictive algorithms that differ from conventional linear prediction in how they take advantage of the previous samples of *s(n)* in the prediction. Conventional linear prediction forms a linear combination from p previous samples by treating each of these as such. In our proposed methods, *kp*+l ($k \in N, l \in Z^+$) samples previous to *s(n)* are first grouped and the values obtained from the groups are then extrapolated at time instant *n* in order to obtain p values to be used as p data samples in the prediction. The prediction is then optimised in the same way as in conventional LP by minimising the square of the prediction error and solving for the normal equations. The number of unknowns in the normal equations will be p, yet the order of the resulting all-pole filter will equal *kp*+l.

The values from the sample groups are obtained by forming a regression line from the samples, for example in Regressive Linear Prediction with Triplets (RLPT) (see P[3]) p groups of three samples (*s(n-2i-1)*, *s(n-2i)* and *s(n-2i+1)*) are formed from 2*p*+1 previous samples of *s(n)*. Every second sample belongs to two triplets. Each triplet defines a regression line, which is then extrapolated at time instant *n*.

Another algorithm, Linear Prediction with Sample Grouping (LPSG) (see P[8]) divides the samples into two main groups: Samples *s(n-1)* and *s(n-2)* form one, a more emphasised, group of samples, which is supposed to have larger correlation with *s(n)*. The p other samples (*s(n-3), s(n-p-2)*) belong to the second group. Each sample in the latter group forms a regression line with the two samples in the first group (i.e. samples *s(n-i-2*), *s(n-1)* and *s(n-2)*). The values of these regression lines at time instant *n* are used as data samples in the prediction.

In this thesis six new predictive algorithms are presented, which all group and extrapolate the past samples prior to the prediction. These operations (i.e., forming regression lines) are linear; hence a linear transform precedes the actual linear combination.

4.2 Experiments

In the papers published, different grouping methods of the past samples of *s(n)* have been analysed. Groups of two or three consecutive samples have been applied for generating new data samples using regression lines. There are also different algorithms depending on whether a certain sample *s(n-i)* is included in one or more groups of samples (Varho and Alku, 1997; Varho and Alku, 1998b; Varho and Alku, 1998c, Alku and Varho, 1998d; Varho and Alku, 1999; Varho and Alku 2000a; Varho and Alku, 2000c) Additionally, two algorithms are presented where either only sample *s(n-1)* or two samples, *s(n-1)* and *s(n-2)*, are additionally emphasised (Alku and Varho, 1997 ; Varho and Alku, 1998a; Varho and Alku 2000b; Varho and Alku; 2000d).

The rationale for the algorithmic development of this study is related to speech coding: we aim at new predictive techniques that would model the speech spectrum more accurately without increasing the number of parameters required for constructing that model. Therefore, we have analysed the behaviour of algorithms

with small to moderate p-values (p equalling 2-12). We have applied telephone bandwidth in our experiments. In addition, the length of the analysis frame was 32 ms, which conforms to typical values in coding applications.

Since in all of these new algorithms the resulting all-pole filter will be of an order larger than p, normally at least one of the filter's poles will be real and located close to the unit circle. Cancellation of the standard pre-emphasiser (Chapter 2.5, $H(z) = 1 + 0.9z^{-1}$) would also yield an extra pole at the real axis, close to the one that the proposed algorithms generate. Therefore, this type of a pre-emphasiser is not optimal for the proposed predictive techniques. However, a first-order IIR filter with its pole at $z=-0.9$ would have a similar effect of reducing spectral dynamics, yet its cancellation would result in a zero on the other side of the x-axis than the pole from the modelling filter. Hence, in some of the algorithms this IIR pre-emphasiser was used instead of the FIR pre-emphasiser that conventional LP applies.

In our experiments both real and synthetic speech was used. The speech material consisted of both female and male speakers pronouncing Finnish words. The main focus of the study was on voiced speech, in particular vowels. However, the modelling of nasals and fricatives was also analysed in our comparison of different predictive methods.

The new all-pole modelling techniques were compared primarily to conventional linear prediction; however, Discrete-All-Pole Modeling (DAP) (El-Jaroudi and Makhoul, 1991) was also used as a reference in our experiments. New predictive methods were assessed by analysing both the performance in modelling a spectral envelope and the corresponding residuals. To measure the modelling of the spectral envelope we used signal-to-error rate (SER) between the spectra of the original voice signal $S(z)$ and the all-pole filter $A(z)$. SER is defined on the logarithmic scale as follows:

$$SER = \frac{\sum_{z=z_1}^{M/2} \left(20\lg|S(k)|\right)^2}{\sum_{z=z_1}^{M/2} \left(20\lg|S(k)| - 20\lg|H(k)|\right)^2} \tag{4.1}$$

where *S(k)* and *H(k)* denote the spectrum of speech and the all-pole filter, respectively. *M* denotes the size of the FFT (i.e., 512), and z_1 is the starting index of the analysis range. The energies of the spectra were normalised to unity.

In addition, the number of formants and their locations were studied. The spectral flatness of the residual spectra was compared using the Itakura-Saito spectral flatness measure (Markel and Gray, 1976):

$$IS = \frac{10}{M/2+1}\left\{\sum_{i=0}^{M/2} \lg|E(i)|\right\} - 10\lg\left\{\frac{\sum_{i=0}^{M/2}|E(i)|}{M/2+1}\right\} \tag{4.2}$$

where *M* denotes the size of FFT (i.e., 512), and *E(i)* denotes the residual energy.

5. Contribution of the Author

This doctoral thesis has been carried out in the Graduate School for Electronics, Telecommunication and Automation (GETA), financed by Finnish Academy. During the years 1996-1999 I have been working at the University of Turku at the Electronics and Information Technology laboratory, and during the year 2000 I have carried out the research at the Helsinki University of Technology at the Acoustics and Audio Signal Processing laboratory.

My supervisor, the creator and my partner in this research has been professor Paavo Alku. He is the father of the basic idea and the developer of the algorithms. In this research group of two members, my responsibility has been the analysis of the behaviour of the algorithms and the optimisation of the parameters. The order of authors in the publication titles indicates that the first author has actually written the paper. However, in all the publications professor Alku has given a very strong impact on the work, from the first idea to the finished paper.

6. Conclusions

Linear predictive coding has been widely studied and applied to speech processing already for decades. The basic idea is to predict sample *s(n)* by forming an optimal linear combination of p previous samples of *s(n)* (Atal and Hanauer, 1971). In the frequency domain, linear prediction corresponds to modelling the spectral envelope by an all-pole filter. In order to achieve better speech quality, i.e., better matching of speech formants, the all-pole filter has to have enough poles to model the local maxima. In speech coding applications the resources for all-pole modelling are limited, i.e., the available number of parameters to define an all-pole filter is moderate, typically p = 8, 10 or 12. This thesis presents six new predictive techniques aiming at more accurate modelling of speech spectra with a compressed set of parameters. In the proposed predictive techniques the order of the all-pole filter has been increased (*kp*+l) while keeping the number of parameters required for characterising the filter equal to p. Lengthening the all-pole filter has been carried out by taking more past samples into prediction. In order not to increase the number of parameters to be transmitted, the past values of the signal have been grouped and extrapolated before the actual linear prediction.

When compared to conventional linear prediction the proposed predictive techniques yield better spectral matching. The new all-pole modelling algorithms

were able to model more formants, and often the proposed method could find the highest formant while conventional LP missed it. The new methods also located formants more accurately, especially the most important low frequencies were more accurate for the new techniques than for conventional LP. Partly due to the different pre-emphasiser, the all-pole spectra given by the new techniques could avoid the so-called low-frequency boost (i.e., Wong et al., 1980). The largest differences appeared in residual energies: compared to LP the new all-pole modelling techniques yielded considerably smaller residual energies in almost all the cases studied.

The results appear very promising: the quality of the speech model has been improved with respect to the number of parameters that define the model. However, the new predictive techniques require somewhat more computation than LP, since the normal equations cannot be presented by means of the Toeplitz matrix. However, most of the computational load in LP comes from the computation of autocorrelation coefficients, and hence the increase in computation is not severe. Another drawback in the proposed methods is that the stability of the resulting all-pole filter cannot be guaranteed. In our experiments though, an unstable filter occurred very seldom, and even in the case of instability there are methods to cope with it: the filter can be stabilised by reflecting the pole outside the unit circle inside the unit circle, or by moving the pole at the unit circle slightly inside, towards the origo.

In this thesis only scalar quantisation was studied with rather small prediction orders. The results of preliminary scalar quantisation seemed promising, yet it would be challenging to design an optimal vector quantisation for the proposed techniques in the future.

References

Alkázar-Fernández, J. and Fraile-Peláez, F. J. 1988. Linear prediction with L_p norm minimization, *Signal Processing IV: Theories and applications*, pp. 1109-1112.

Alku, P. 1992. Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering, *Speech Communication*, **11**(2-3): 109-118.

Alku, P. and Varho, S. 1997. A new linear predictive method for spectral estimation of voiced speech, *Proceedings of 1997 IEEE International Symposium on Circuits and Systems*, Hong Kong, Vol. IV, pp. 2649-2652.

Atal, B. S. and Hanauer, S. L. 1971. Speech analysis and synthesis by linear prediction of the speech wave, *Journal of the Acoustical Society of America*, **50**(2): 637-655.

Atal, B. S. and Schroeder, M. S. 1979. Predictive coding of speech signals and subjective error criteria, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **27**(3): 247-254.

Baken, R. J. and Orlikoff, R. F. 1999. *Clinical Measurement of Speech & Voice*, Delmar Publishers, 2nd edition.

Baselli, G. et al. 1985. Autoregressive modeling and power spectral estimate of R-R interval time series in arrhythmic patients, *Computers and Biomedical Research*, 18: 510-530.

Bimbot, F., Blomberg, M., Boves, D., Genoud, D., Hutter, H.-P., Jaboulet, C. Koolwaaij, J., Lindberg, J. and Pierrot, J.-B. 2000. An overview of the CAVE project research activities in speaker verification, *Speech Communication*, **31**(2-3): 155-180.

Birgmeier, M., Bernhard, H.-P. and Kubin, G. 1997. Nonlinear long-term prediction of speech signals. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1283-1286.

Bourget, C., Aboulnasr, T. and Verreault, E. 1995. Perceptual speech coding, *Canadian Conference on Electrical Engineering*, pp. 1070-1072.

Cheung, Y., Lai, Z. and Xu, L. 1996. Application of adaptive RPCL-CLP with trading System to foreign exchange investment. *Proceedings of IEEE International Conference on Neural Networks*, New York, NY, Vol. 4, pp. 2033-2038.

Choi, S. H., Kim, K. K. and Lee, H. S. 2000. Speech recognition using quantized LSP parameters and their transformations in digital communication, Speech Communication, 30(4): 223-233.

Chou, S., Chen, C., Yang, C. and Lai, F. 1996. A rule-based neural stock trading decision support system, *Proceedings of the Conference on Computational Intelligence for Financial Engineering*, New York, NY, pp. 148-154.

David, S. and Ramamurthi, B. 1991. Two-sided filters for frame-based prediction, *IEEE Transactions on Signal Processing*, **39**(4): 789-794.

Deller, J. R., Proakis, J. G. and Hansen, J. H. L. 1993. *Discrete-Time Processing of Speech Signals*, Prentice-Hall, Inc., Englewood Cliffs, NJ.

Deneire, L. and Slock, D. T. M. 1999. A Schur method for multiuser multichannel blind identification, *IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 5, pp. 2905-2908.

Denoël, E. and Solvay, J-P. 1985. Linear prediction of speech with a least absolute error criterion, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **33**(6): 1397-1403.

Dinda, P. A. and O'Hallaron, D. R. 1999. An evaluation of linear models for host load prediction, *Proceedings of the Eighth International Symposium on High Performance Distributed Computing*, pp. 87-96.

El-Jaroudi, A. and Makhoul, J. 1991. Discrete all-pole modeling, *IEEE Transactions on Signal Processing*, **39**(2): 411-423.

ETSI 1997. ETSI 300 961. Digital cellular telecommunications system; Full rate speech; Transcoding (GSM 06.10 version 5.0.1), Sophia Antipolis.

Fant, G. C. M. 1960. *Acoustic Theory of Speech Production*, Mouton, Gravenhage, The Netherlands.

Fenwick, P. B. C., Michie, P., Dollimore, J. and Fenton, G. W. 1971. Mathematical simulation of the electroencephalogram using an autoregressive series, *Biomedical Computing*, **2**: 281-307.

Flanagan, J. L., Schroeder, M.R., Atal, B. S., Crochier, R. E., Jayant, N. S., and Tribolet, J. M. 1979. Speech coding, *IEEE Transactions on Communications*, **27**(4): 710-737.

Gersho, A. 1994. Advances in speech and audio compression, *Proceedings of IEEE*, **82**(6): 900-918.

Golchin, F. and Paliwal, K. P. 1997. Classified adaptive prediction and entropy coding for lossless coding of images, *Proceedings of the International Conference on Image Processing*,. pp. 110-113.

Gray, A. H., Gray, R. M. and Markel, J. D. 1977. Comparison of optimal quantizations of speech reflection coefficients, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **25**: 9-23.

Hanson, H., Maragos, P. and Potamianos, A. 1994. A system for finding speech formants and modulations via energy separation, *IEEE Transactions on Speech and Audio Processing*, **2**(3): 436-443.

Hermansky, H. 1990. Perceptual linear predictive (PLP) analysis of speech, *Journal of the Acoustical Society of America*, **87**(4): 1738-1752.

Hsue, J. J. and Yagle, A. E. 1993. Fast algorithms for close-to-Toepliz-plus-Hankel systems and two-sided linear prediction, *IEEE Transactions on Signal Processing*, **41**(7): 2349-2361.

Höntsch, I. and Karam, L. J. 1997. APIC: Adaptive perceptual image coding based on subband decomposition with locally adaptive perceptual weighting, *Proceedings of the IEEE International Conference on Image Processing*, pp. 37-40.

Kay, S. M. 1983. Some results in linear interpolation theory, *IEEE Transactions on Acoustical Speech Signal Processing*, **31**(3): 746-749.

Kay, S. M. 1988. *Modern Spectral Estimation: Theory and Application*, Prentice-Hall, Inc., Englewood Cliffs, NJ.

Kakusho, O., Yanagida, M. and Mizogushi, R. 1984. A sample selective linear prediction analysis of speech, *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, 1984, paper 2.1.

Kleijn, W.B. and Paliwal, K.K. 1995. *Speech Coding and Synthesis*, Elsevier Science, Amsterdam, The Netherlands.

Kondoz, A. M., 1995. *Digital Speech Coding for Low Bit Rate Communication Systems*, Wiley, New York.

Kroon, P. and Atal, B. S. 1987. Quantization procedures for excitation in CELP coders, *Proceedings of International Conference on Acoustics and Speech Signal Processing*, pp. 1649-1652.

Laine, U. K. 1995. Generalized linear prediction based on analytic signals, *Proceedings of IEEE International Conference on Acoustics and Speech Signal Processing*, pp. 1701-1704.

Lee, C. 1988. On robust linear prediction, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **30**(5): 642-650.

Lee, A. C. 1989. A new autoregressive method for high-performance spectrum analysis, *Journal of the Acoustical Society of America*, **86**:(150-157).

Lee, W. S. 1999. Edge adaptive prediction for lossless image coding, *Proceedings of Data Compression Conference,* Session 9: 6/8, pp. 1-8.

Ma, C., Kamp, Y. and Willems, L. F. 1993. Robust signal selection for linear prediction analysis of voiced speech, *Speech Communication*, Vol. 12, pp. 69-81.

Makhoul, J. 1975. Linear prediction: A tutorial review, *Proceedings of IEEE*, **63**(4): 561-580.

Marchand, J. F. P. and Rhody, H. E. 1997. Noncausal image prediction and reconstruction, *Proceedings of Data Compression Conference,* Poster session, p. 453.

Markel, J. D. and Gray, A. H. Jr. 1976. *Linear Prediction of Speech*, Springer, New York.

Marple, S. L. Jr. 1982. Fast algorithms for linear prediction and system identification filters with linear phase, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **30**(6): 942-953.

Marple, S. L. Jr. 1987. *Digital Spectral Analysis*, Prentice-Hall International.

Miyoshi, Y., Yamato, K., Mizoguchi, R., Yanagida, M. and Kakusho, O. 1987. Analysis of speech signals of short pitch period by a sample-selective linear prediction, *IEEE Transactions on Acoustics and Speech Signal Processing*, **35**(9): 1233-1240.

Motta, G., Storer, J. A. and Carpentieri, B. 1999. Adaptive linear prediction lossless image coding, *Proceedings of Data Compression Conference*, pp. 491-500.

Nordling, C. and Österman, J. 1980. *Physics Handbook*, Bratt Institut für Neues Lernen, Lund.

Oppenheim, A. V. and Schafer, R. W. 1989. *Discrete-Time Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ.

Parsons, T. 1986. *Voice and Speech Processing*, McGraw-Hill College Div., Inc.

Rabiner, L.R. and Schafer, R.W. 1978. *Digital Processing of Speech Signals*, Prentice-Hall, Inc., Englewood Cliffs, NJ.

Reddy, N. S. and Swamy, M. N. S. 1984. High-resolution formant extraction from linear-prediction phase spectra, *IEEE Transactions on Acoustics, Speech and Signal Processing*, **32**(6): 1136-1144.

Schroeder, M. R. and Atal, A. S. 1985. Code-excited linear prediction (CELP): High-quality speech at very low bit rates, *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 937-940.

Shimamura, T. and Hayakawa, H. 1999. Adaptive nonlinear prediction based on order statistics for speech signals, *Proceedings of 6th European Conference on Speech Communication and Technology*, Budapest, pp. 347-350.

Steiglitz, K. and Dickinson, B. 1977. The use of time-domain selection for improved linear prediction, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **25**(1): 34-39.

Strube, H. W. 1980. Linear prediction on a warped frequency scale, *Journal of the Acoustical Society of America*, **68**(4): 1071-1076.

Townshend, B. 1991. Nonlinear prediction of speech, *Proceedings of IEEE International Conference Acoustics and Speech Signal Processing*, pp. 425-428.

Varho, S. and Alku, P. 2000. All-pole spectral modelling of voiced speech with a highly compressed set of parameters, *Proceedings of X European Signal Processing Conference*, Tampere, Finland, Vol. I, p. 59.

Viswanathan, R. and Makhoul, J. 1975. Quantization Properties of Transmission Parameters in Linear Predictive Systems, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **23**(3): 309-321.

Walker, G. 1931. On periodicity in series of related terms, Proceedings of Royal Society, **131-A**, 518.

Wang, S., Paksoy, E. and Gersho, A. 1994. Performance of nonlinear prediction of speech, *Proceedings of the 3rd International Conference on Spoken Language Processing*, Tokyo, Japan, pp. 29-30.

Weigend, A. S. and Gershenfield, N. 1994. *Time series prediction: Forecasting the future and understanding the past*, Addison-Wesley.

Wong, D. Y., Hsiao, C. C. and Markel, J. D. 1980. Spectral mismatch due to preemphasis in LPC analysis/synthesis, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **28**(2): 263-264.

Yule, G.U. 1927. On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers, Philosophical Transactions of Royal Society, **226**(A): 267-298.

Öztürk, Y. and Abut, H. 1992. Multichannel Multidimensional Linear Predictive Systems, IEEE Transactions on Image Processing, **IP-1**(1): 101-106.

Appendix A: Publications P1 – P9

Publication P1

Varho, S. and Alku, P. 1997. A Linear Predictive Method Using Extrapolated Samples for Modelling of Voiced Speech, *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, Session IV, pp. 13-16.

Publication P2

Varho, S. and Alku, P. 1998a. Separated Linear Prediction - A new all-pole modelling technique for speech analysis, *Speech Communication,* **24**: 111-121.

Publication P3

Varho, S. and Alku, P. 1998b. Regressive Linear Prediction with Triplets - An Effective All-Pole Modelling Technique for Speech Processing, *Proceedings of IEEE International Symposium on Circuits and Systems*, Monterey, CA, Vol. IV, pp. 194-197.

Publication P4

Varho, S. and Alku, P. 1998c. Spectral Estimation of Voiced Speech with Regressive Linear Prediction, *Proceedings of IX European Signal Processing Conference*, Rhodes, Greece, Vol. II, pp. 1189-1192.

Publication P5

Alku, P. and Varho, S. 1998d. A New Linear Predictive Method for Compression of Speech Signals, *Proceedings of the 5th International Conference on Spoken Language Processing*, Sydney, Australia, Vol. VI, pp. 2563-2566.

Publication P6

Varho, S. and Alku, P. 1999. A New Predictive Method for All-Pole Modelling of Speech Spectra with a Compressed Set of Parameters, *Proceedings of IEEE International Symposium on Circuits and Systems*, Orlando, FL, Vol. III, pp. 126-129.

Publication P7

Varho, S. and Alku, P. 2000a. A Linear Predictive Method for Highly Compressed Presentation of Speech Spectra, *Proceedings of IEEE International Symposium on Circuits and Systems*, Geneva, Switzerland, Vol. V, pp. 57-60.

Publication P8

Varho, S. and Alku, P. 2000b. Linear Prediction of Speech by Sample Grouping, *Proceedings of IEEE Nordic Signal Processing Symposium*, Kolmården, Sweden, pp. 113-116.

Publication P9

Varho, S. and Alku, P. 2000d. Separated Linear Prediction - Improved Spectral Modelling by Sample Grouping, *to be published in Proceedings of IEEE International Symposium on Intelligent Signal Processing and Communication Systems*, Honolulu, HI, pp. 731-735.

HELSINKI UNIVERSITY OF TECHNOLOGY
LABORATORY OF ACOUSTICS AND AUDIO SIGNAL PROCESSING

33 P. Alku: An Automatic Inverse Filtering Method for the Analysis of Glottal Waveforms. 1992

34 V. Välimäki: Fractional Delay Waveguide Modeling of Acoustic Tubes. 1994

35 T. I. Laakso, V. Välimäki, M. Karjalainen, U. K. Laine: Crushing the Delay—Tools for Fractional Delay Filter Design. 1994

36 J. Backman, J. Huopaniemi, M. Rahkila (toim.): Tilakuuleminen ja auralisaatio. Akustiikan seminaari 1995

37 V. Välimäki: Discrete-Time Modeling of Acoustic Tubes Using Fractional Delay Filters. 1995

38 T. Lahti: Akustinen mittaustekniikka. 2. korjattu painos. 1997

39 M. Karjalainen, V. Välimäki (toim.): Akustisten järjestelmien diskreettiaikaiset mallit ja soittimien mallipohjainen äänisynteesi. Äänenkäsittelyn seminaari 1995

40 M. Karjalainen (toim.): Aktiivinen äänenhallinta. Akustiikan seminaari 1996

41 M. Karjalainen (toim.): Digitaaliaudion signaalinkäsittelymenetelmiä. Äänenkäsittelyn seminaari 1996

42 M. Huotilainen, J. Sinkkonen, H. Tiitinen, R. J. Ilmoniemi, E. Pekkonen, L. Parkkonen, R. Näätänen: Intensity Representation in the Human Auditory Cortex. 1997

43 M. Huotilainen: Magnetoencephalography in the Study of Cortical Auditory Processing. 1997

44 M. Karjalainen, J. Backman, L. Savioja (toim.): Akustiikan laskennallinen mallintaminen. Akustiikan seminaari 1997

45 V. Välimäki, M. Karjalainen (toim.): Aktiivisen melunvaimennuksen signaalinkäsittelyalgoritmit. Äänenkäsittelyn seminaari 1997

46 T. Tolonen: Model-Based Analysis and Resynthesis of Acoustic Guitar Tones. 1998

47 H. Järveläinen, M. Karjalainen, P. Maijala, K. Saarinen, J. Tanttari: Työkoneiden ohjaamomelun häiritsevyys ja sen vähentäminen. 1998

48 T. Tolonen, V. Välimäki, M. Karjalainen: Evaluation of Modern Sound Synthesis Methods. 1998

49 M. Karjalainen, V. Välimäki (toim.): Äänenlaatu. Akustiikan seminaari 1998

50 V. Välimäki , M. Karjalainen (toim.): Signaalinkäsittely audiotekniikassa, akustiikassa musiikissa. Äänenkäsittelyn seminaari 1998

51 M. Karjalainen: Kommunikaatioakustiikka. 1998

52 M. Karjalainen (toim.): Kuulon mallit ja niiden sovellutukset. Akustiikan seminaari 1999

53 Huopaniemi, Jyri: Virtual Acoustics And 3-D Sound In Multimedia Signal Processing. 1999

54 Bank, Balázs: Physics-Based Sound Synthesis of the Piano. 2000

55 Tolonen, Tero: Object-Based Sound Source Modeling. 2000

56 Hongisto, Valtteri: Airborne Sound Insulation of Wall Structures — Measurement And Prediction Methods. 2000

57 Zacharov, Nick: Perceptual Studies On Spatial Sound Reproduction Systems. 2000

ISBN 951-22-5409-3

ISSN 1456-6303

Report 58 Susanna Varho NEW LINEAR PREDICTIVE METHODS FOR DIGITAL SPEECH PROCESSING