MASTER

Stereo coding by two-channel linear prediction and rotation

Selten, T.P.J.

Award date:
2004

Link to publication

Disclaimer
This document contains a student thesis (bachelor's or master's), as authored by a student at Eindhoven University of Technology. Student theses are made available in the TU/e repository upon obtaining the required degree. The grade received is not published on the document as presented in the repository. The required complexity or quality of research of student theses may vary by program, and the required minimum study period may vary in duration.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
Stereo coding by
two-channel linear
prediction and rotation

by T.P.J. Selten

Master of Science thesis

Project period: 09-2003 – 10-2004
Report Number: 19-04

Commissioned by:

Supervisors:
Dr.ir. A.C. den Brinker (Philips Research)
Dr.ir. L.L.M. Vogten (TU/e)

Additional Commission members:
Prof.dr.ir. J.W.M. Bergmans (TU/e)

The Department of Electrical Engineering of the Eindhoven University of Technology accepts no responsibility for the contents of M.Sc. theses or practical training reports
TECHNISCHE UNIVERSITEIT EINDHOVEN

Department of Electrical Engineering

MASTER'S THESIS

Stereo coding by two-channel linear prediction and rotation

by

T.P.J. Selten

Supervisors : prof. dr. ir. J.W.M. Bergmans (TU/e)
              : dr. ir. L.L.M. Vogten (TU/e)
              : dr. ir. A.C. den Brinker (Philips Research)

Eindhoven, 20th October 2004
"If it will take ten years to make the machine with available technology, and only five years to make it with a new technology, and it will only take two years to invent the new technology, then you can do it in seven years by inventing the new technology first!"

Neal Stephenson in Cryptonomicon, 1999
Summary

In the context of an M.Sc. graduation project at Philips Research Laboratories in Eindhoven, this thesis concerns stereo coding as part of an audio coder. Audio coding aims at removing redundancies and irrelevancies from the audio signal to reduce the bit rate. Stereo coding aims at exploiting cross-channel redundancies and irrelevancies to attain a lower bit rate than the sum of the independently coded channels.

Desired is a stereo coding technique which allows perfect reconstruction at the high bit rate end (e.g., in the absence of quantization), and which is able to compete with the latest stereo coding methods at the low bit rates.

Known stereo coding methods, briefly discussed in this thesis, do not meet the desired criteria. The most efficient technique is parametric stereo coding (Binaural Cue Coding (BCC) and Optimal Coding of Stereo (OSC)) which separates the signal into audio content and spatial information. By reducing the two audio channels to one and some side information, very low bit rates are achieved. However, the original can not be reconstructed perfectly.

In this thesis, we propose to use a two-channel (stereo) Linear Predictor (LP), based on FIR or IIR (Laguerre) filters in combination with a single rotator. A linear estimate of the current signal is made from the history of both channels. The LP allows perfect reconstruction. For the prediction filters Laguerre filters are chosen, because of their close resemblance to the psycho-acoustical Bark-scale, which is advantageous for lossy audio coding.

Optimizing the proposed system dissolves in two separate stages: optimizing the stereo LP and subsequently optimizing the rotator with the residual from the Stereo LP. An efficient way of optimizing the Stereo LP is using the block-Levinson algorithm, which only applies under the condition of equal orders of auto and cross predictors. Furthermore, two preliminary ideas for quantization of these optimal parameters are introduced in this thesis.

Stability of the proposed system is examined more closely in this thesis, the stability being determined through the synthesis filter of the Stereo LP. It is argued that stability is only guaranteed when equal orders of auto and cross predictors are used.

The proposed system is implemented in Matlab in combination with C, and works as expected. Preliminary experiments have been conducted to establish which parts of the residual signal have to be maintained in a transmission system and which parts can be compromised. Unlike the situation in BCC and OCS, the side signal can apparently not be discarded completely without causing problems, but fair results are already achieved by only transmitting 10% of the side signal, i.e. only low frequencies. Other methods are examined that would allow low bit rate coding and the most promising method appears to be Spectral Band Replication (SBR). This technique is very suitable due to the spectrally flat character of residual signals. With SBR only a part of the total bandwidth is transmitted, and this is applicable to both the main and the side signal. This suggests that low bit rate coders based on the proposed system are feasible.
Preface

The thesis before you is the result of my nine-months' graduation project, which has been conducted at Philips Research in order to obtain the degree Master of Science in Electrical Engineering from the University of Technology in Eindhoven (TU/e). This project has been initiated by the Audio Coding cluster, which is part of the Digital Signal Processing (DSP) group. This project was accepted under the responsibility of the research chair Signal Processing Systems, which is part of the capacity group Measurement and Control Systems at the department of Electrical Engineering of TU/e.

Through this project I've seen, and participated in a larger research project aimed at developing audio coding algorithms. Thus I got involved in different stages of that project ranging from the stage of an early idea-like, the starting point of this project, until extensive listening tests for finished coders. It also let me "re-discover" music, which was gradually been pushed to the background. Attentive listening sometimes required for this project rubbed of to "normal" music listening, intensifying the experience, with maybe the exception of some excerpts of music and speech used for testing.

I would like to acknowledge the institutions and individuals that made it possible to deliver this thesis. First of all Philips Research Eindhoven which provided an inspiring working environment for this thesis, and furthermore I want to express my gratitude to my supervisors: dr. ir. Bert den Brinker at Philips Research, and to dr. ir. Leo Vogten and prof. dr. ir. Jan Bergmans for their guidance and support. I especially want to thank Bert for helping me out with writing this thesis. Furthermore, I want to thank my colleagues at the DSP group, and particularly Arijit Biswas M.Sc. for his close cooperation with this project. And finally, I would like to thank all the students who inhabited the DSP-student room during my time at Philips, making it such a pleasant stay.
# Contents

Summary                                    i  
Preface                                     iii

1. Introduction  
   1.1. Problem description and statement ........................................... 3  
   1.2. Outline of this report ............................................................. 4

2. Background  
   2.1. Binaural perception and stereo signals .......................................... 5  
   2.2. Stereo Coding  
      2.2.1. Sum-Difference coding ............................................................ 8  
      2.2.2. Intensity stereo coding .......................................................... 10  
      2.2.3. Stereo linear prediction ....................................................... 11  
      2.2.4. Complex linear prediction ..................................................... 13  
      2.2.5. Rotation (VECMA) ................................................................. 13  
      2.2.6. Parametric stereo coding ...................................................... 14

3. Stereo LPC  
   3.1. Two-channel LP and rotation .......................................................... 18  
   3.2. Prediction filters ................................................................. 20  
   3.3. Calculation of the optimal parameters ........................................... 22  
   3.4. Practical problems and solutions ................................................ 26  
      3.4.1. Stability of the 2-channel synthesis filter ................................ 27  
      3.4.2. Regularization of the Stereo LP optimization problem .................... 28  
      3.4.3. Robustness of the stereo LP system ......................................... 28  
      3.4.4. Regularization of the rotator optimization .................................. 30

4. Quantization of the Stereo LPC coefficients  
   4.1. Transmission methods for one-channel Linear Prediction ...................... 31  
   4.2. Quantization of two-channel prediction parameters  
      4.2.1. Reflection Matrices ............................................................... 34  
      4.2.2. Coupling to OCS parameters .................................................. 34
# Contents

5. **Signal quantization experiments**  
   5.1. Signals in the system ........................................ 37  
   5.2. Uniform quantization ........................................ 40  
   5.3. Nonuniform quantization ..................................... 41  
   5.4. Side signal removal .......................................... 42  
   5.5. Synthetic side signal ........................................ 43  
   5.6. Partial side signal transmission ............................ 43  
   5.7. Spectral bandwidth replication .............................. 44  
   5.8. Conclusion from the experiments ............................ 45  

6. **Conclusions and Recommendations**  
   6.1. Conclusions .................................................. 47  
   6.2. Recommendations ............................................. 48  

A. **Lists of abbreviations, notations and used excerpts**  
   A.1. List of abbreviations ........................................ 49  
   A.2. List of notations ............................................. 50  
   A.3. List of used excerpts ....................................... 50  

B. **One-channel linear prediction** ................................ 51  

C. **Rotation** ..................................................... 57  

D. **Relation between Laguerre and tapped-delay-line LP systems**  
   ................................................................. 61  

E. **Separability of optimization** ................................ 63  

F. **Matlab & C headers** ......................................... 67  

**Bibliography** ..................................................... 73
1. Introduction

When listening to a live music performance, one can perceive the surroundings just by listening. You are able to pinpoint and track various sound sources on the stage, get a feeling of the size of the room played in (e.g. when played in a church) etc.. This because we are listening with a pair of ears. The spatial perception depends solely on the brain’s processing of the acoustical signals received by both ears, and are condensed in a limited number of cues. These acoustical cues are in the form of intensity and time differences between both ears, but also interaural coherence plays a role. With these different cues we are possible to localize sound sources in a 3 dimensional space, with a considerably high accuracy (but dependent on the content of the sound and its source position).

But these spatial cues are no longer present in a monophonic reproduction. Hence, stereophonic reproduction and of course recording methods were developed for trying to recreate the spatial sound image, similar to the live listening experience. The term stereophonic is derived from two Greek words: stereo, which means solid and implies multiple dimensions, and phonics, which means the science of sound. In the 1930’s, Blumlein from Thorn E.M.I. and Fletcher et al. from Bell Laboratories independently did most of pioneering work for stereophonic recording and reproduction techniques [14]. For the reproduction of these early stereophonic sound multiple loudspeakers were used, ranging from 2 up to 80 (Fletcher’s wall of sound used an array of 80 microphones each connected to a corresponding loudspeaker, placed in an identical position, in a separate room). The movie industry applied the stereo sound in the 1960, and pushed the multichannel sound into the home market. Stereo is currently a widely accepted contraction of stereophonic, which generally refers to two-channel systems (see Fig. 1.1). Also headphones were used for reproduction and this is then called binaural or biphonic listening. An advantage of using headphones is that it is easier to control the received sound for the ears independently; with loudspeakers this left-right integrity of sound is not guaranteed.

The stereo sound was introduced and matured during the analog era, but since the introduction of the Compact Disc (CD) in the 1980’s the recording and reproduction of wide-band audio has been shifted from the analog to the digital form. Nowadays the presence of the CD is ubiquitous. This digital form has quite some advantages over the analog way, like: high fidelity audio, a big dynamic range and its robustness. To represent the audio in digital format, quantization (subdivision into small but measurable increments for discrete moments in time) is needed and thus noise is introduced. To make the quantization noise inaudible, a sample frequency of 44.1 kHz was chosen at a resolution of 16 bits per sample which are coded with PCM. Because the CD mostly contains stereo audio, this results in a total bit rate of $44100 \times 16 \times 2 = 1411200$ bits/sec. For some applications this high bit rate is an obstacle. The audio quality of the CD is considered to be transparent which means that the reconstructed audio output cannot be
In spite of the fact that storage and transmission media are getting bigger and faster, there are still lots of audio and speech applications with limited bandwidth or storage capacities which also call for high quality audio. So the demand for efficient storage or transmission of data still exists, like e.g. for audio and speech applications in network or mobile systems. Luckily, audio and speech data contains excess information which can be removed without loss in quality and which makes it possible to reduce the data rate.

The encoding and decoding algorithms (codecs) for compressing audio data have several attributes, like: bit rate, quality, delay, computational complexity, robustness, scalability etc.. These can be traded off depending on the application.

There are two main principles which coding algorithms can exploit for bit rate reduction. The first one exploits redundancies within the signal (signal redundancies), an audio signal may be partly predictable from its past or can be represented in a more efficient way e.g. describing the waveform as a combination of sinusoids which are defined by amplitude, frequency and starting phase. The second one is perceptual irrelevancies which makes use of the properties of the human auditory system. An audio signal usually contains components which can be removed without a perceptual loss in quality. Making use of auditory masking effects is an example of a perceptual irrelevancy, e.g. a signal can be inaudible during a simultaneously occurring stronger signal.

The coding algorithms can consequently be divided into two categories namely: Lossless and Lossy coders. The lossless coders can reconstruct the original waveform perfectly from the coded representation. There is of course no loss in quality because the waveform is exactly the same. Linear Prediction (LP) can be used as a lossless technique because LP is in the absence of quantization a reversible process. These lossless coders can only exploit signal redundancies and can achieve a gain reduction of about 50-60%, but this is highly signal dependent.

In contrast to the lossless coders, a lossy coder can benefit from both redundancies and irrelevancies and can therefore achieve higher bit rate reductions. One of the most widespread
lossy coding standards for digital audio encoding is the MPEG-1 layer III standard, generally referred to as MP3. The achieved gain of these lossy coders is about 90%, e.g., MP3 can reduce a PCM file with a size of 40 megabyte to a 4 megabyte file. This makes them useful for network based applications. Lossy coders mainly depend on a psychoacoustic model which is based on knowledge of the human auditory system.

Since audio signals usually consist of at least two channels, which are supposed to be often related in some way, it may also be worthwhile to make use of inter-channel redundancies and irrelevancies. In [1] however is stated that there is not much correlation between the channels in the (short-term) time domain. But it can be easily shown that when recording a single sound source with a simple stereo microphone, the channels are related by their transfer functions. The cross-correlation in the frequency domain, therefore, shows more dependencies, especially the magnitudes of the spectral coefficients [2]. However with the current digital mixing methods it is also easily feasible to create an artificial or synthetic stereo audio signal between which no cross-channel correlation is guaranteed.

There are already techniques which exploit cross-channel redundancies and/or irrelevancies, and these will be discussed later on in this report. But these techniques do not meet the desired criteria, hence this project was started.

1.1. Problem description and statement

This graduation project is done within the coding cluster of the DSP group from Philips Research Laboratories Eindhoven. The context of this is project is stereo coding. Stereo coding aims at removing redundancy and irrelevancy from the stereo signal to attain lower bit rates than the sum of the bit rates of the independently coded channels while maintaining the same quality level. Some known stereo coding techniques, which are briefly described in section 2.2, are: mid/side stereo coding, intensity coding, rotators over the total band or per frequency band, and parametric stereo coding. The most promising technique appears to be parametric stereo coding, it outperforms standard solutions like mid/side and intensity stereo coding. A problem with this coding is that the original signal can not be reconstructed perfectly due to the used perceptual model.

This brings us to the problem statement of this thesis: We desire a stereo coding technique subject to the following criteria:

1. Encoder and decoder form a system allowing perfect signal reconstructing in the absence of signal quantization and, thus, near perfect signal reconstruction at the high bit rate end.

2. The encoder constructs a main and a side signal similar to those provided by parametric perceptual stereo coding since this is advantageous for low bit rate coding purposes.

We propose to use a technique which is able to meet these constrains namely Linear Predictive Coding (LPC), based on FIR or IIR-filtering, in combination with a single rotator. LPC is able to reduce the redundancy in stereo signals and it is able to reconstruct the original perfectly, in the absence of quantization. It is also an often employed tool in audio and speech coding.
but not widely used for stereo coding. In the proposed system LPC is combined with a single rotator for constructing the main and the side signal, like in parametric stereo coding.

The objective of this thesis is: to see if it is feasible to extend or adapt known linear prediction techniques, as already used in speech and audio coding, to a two-channel system, and pinpointing possible problem areas. Furthermore, a very low bit rate is required to successfully compete with the current state-of-the-art stereo audio coders. Therefore experiments have to be performed to examine if low bit rates are achievable with the proposed system. There is even doubt [30] whether using LP is an efficient tool for stereo coding purposes.

1.2. Outline of this report

Chapter 2 provides some background information: a brief introduction into binaural perception and the origin of stereo signals. It furthermore describes the known stereo coding methods with their aims and shortcomings. Chapter 3 introduces the proposed Stereo Linear Prediction scheme with two different prediction filter implementations. In addition, methods are elaborated for optimizing the parameters of the proposed scheme. Finally some problem areas are stated with possible solutions. In chapter 4 existing one-channel LPC quantization schemes are described. Furthermore two preliminary ideas are proposed for quantization of the transmission coefficients. In chapter 5 an example is given to see if the results are in line with the expectations, and also some preliminary experiments which describe possible methods for bit rate reduction. Finally, in chapter 6 conclusions are drawn and some recommendation for future work are made.
2. Background

This chapter provides the reader with some background information desired for this thesis. The reader is assumed to be acquainted with linear prediction and rotation. If this is not the case an explanation can be found in appendix B and C.

This chapter, firstly, describes human sound perception, in particular the binaural perception, and the origin of the stereo signals. Secondly, known stereo coding methods are described detailing why they do not meet our demands.

2.1. Binaural perception and stereo signals

Binaural perception

*Sound* is the perception of air pressure variations picked up through our ears, and by using both ears (binaural) we are able to localize sound sources in 3 dimensional space. This perception of spatial sound(s) depends on the processing of spatial cues. The received sound waves travelled (slightly) different paths to both ears, due to the shape of the head and the position of the ears on the head. Therefore, the sound pressures reaching the ears are not entirely identical. From these differences, spatial cues can be extracted. Cues for spatial hearing can originate from:

- Intensity differences
- Time differences
- Interaural coherence
- Pinnea
- Head/Source Movement

The Interaural Intensity (or Level) Difference (IID) is related to the ratio of intensity of the sound pressures at both ears, the differences in intensity are mainly caused by the head. Sound coming from an off-centered source can vary in intensity in both ears due to the fact that the human head is an obstacle for the sound waves, therefore casts an acoustical shadow (Fig. 2.1), and thus the far ear receives less energy, the difference however is frequency and position dependent. At low frequencies IIDs are insignificant, since the wavelength is much larger than the size of the head, and therefore these low frequencies do not cause large intensity differences (3dB for 500 Hz; 20dB for 6kHz). IIDs due to propagational decay are usually less important than the decrease caused by the acoustic shadow, therefore only considerable for the more distant sound sources.
2.1. Binaural perception and stereo signals

For the same off-centered source, also the direct path for both ears differs in length, which results in a difference in time arrival. This is referred to Interaural Time Difference (ITD). For pure tones, this can be also be expressed as a phase difference (IPD). ITDs are dominant for frequencies below 1500 Hz. The ear which picks up the sound first determines the perceived direction of the sound (law of first wavefront or precedence effect). At the somewhat higher frequencies the human auditory system is not able to detect the fine-structure, however differences in the envelopes of the waveform can still be detected. For the pure tones, the maximum delay approaches at a phase lag of 180°, then the phase difference becomes ambiguous.

ITDs are not unique, i.e. multiple points in space have the same distances to both ears, thus giving the same ITD, and therefore causing a cone of confusion. Fortunately, other cues can resolve this ambiguity.

Not only the differences between both ears give spatial cues but also through similarities cues are obtained. Interaural Coherence (IC), defined as the overall similarity, can capture ambience properties (spatial diffuseness).

By using ITDs and IIDs we can accurately localize sources in the horizontal plane (3.6° in front and 10° to the side [6]), but poorly in the median (vertical) plane. The sound perception in the median plane (vertical) differs because no time or intensity cues can aid in perception. However, changes in the spectrum are used, introduced through diffractions of pinnae and reflections from shoulder and torso. These are Head Related Transfer Functions, e.g. the pinnae can be seen as directional filter.

The previous cues are all static, i.e., it is assumed that listener and source have a fixed position in space. In practice, dynamic changes through head and/or source movement improve our source localizing (or tracking in case of source movement) capabilities and also resolve front-back ambiguity.

The human perceptive system is visual dominant, therefore sources can seem to emanate from a visual source, e.g., lips of a TV actor, rather than from the actual sound source, this is the so-called ventriloquist effect.

Figure 2.1.: Acoustical shadow of the head at high frequencies.
**Stereophonic signals**

The definition of stereophonic from the dictionary (Merriam-Webster) is: relating to, or constituting sound reproduction involving the use of separated microphones and two transmission channels to achieve the sound separation of a live hearing. In other words stereophonic is a way of recording and reproducing the sound will maintaining the complete spatial fidelity. However, Richard Heyser put this in a proper perspective by stating: "Stereo is merely an attempt to create the illusion of reality through the willing suspension of disbelief." [14].

For stereo reproduction, recording is needed and perhaps some processing on these recordings, this to create the two channels. These recordings can originate from a live performance. For capturing this aural experience, multiple microphones are used, placed in specific positions in the performance room. The easiest configuration, for stereo reproduction is just using two microphones (one for each channel), see Fig. 2.2.

![Figure 2.2.: Example of a stereo microphone setup.](image)

For a centered source, equidistant from both microphones, this results in signals with the same intensity and timing characteristics at both microphones. During the reproduction this results in a phantom source placed between the loudspeakers, provided of course that a correct playback system is used. For the off-centered source however, slightly different signals are detected, resulting in inter-channel time and level differences (different path lengths). During playback this results into corresponding cues namely ITDs and IIDs, through this the phantom source is perceived closer to one loudspeaker. However the phantom source position can not be placed outside the loudspeaker arc, it is restrained by the arrangement of the loudspeakers.

For this situation of recording it is clear that there are correlations between the channels, specifically the room transfers from the source to the individual microphones.
Another method for generating a stereo signal is by mixing pre-recorded (mono) signals. These can artificially be placed at different positions within the stereo image, by manipulating timing, intensity, reverberation etc. for both channels individually. For instance, a mono source can be panned by changing the intensity for both channels in a log and reverse-log manner [14]. A panoramic potentiometer (or panpot in short) can be used for this artificially placing of different sources at different positions within the stereo image. By also introducing some timing differences a more natural spatial fidelity can be generated. This is of course a simple method for generating stereo signals, and a clear relation between channels exists.

However, practically mixing is more complex process where all kinds of technical considerations have to be taken into account, i.e which types of microphones to use, the placements of them, what is the goal for the reproduction: is it desired to re-create an sonic event or create a completely new one. Furthermore, different listener’s perspectives are possible [14], i.e. "You are there" and "They are here". With "You are there" the listener is transported into the same sonic environment as the event (the listener is put into the audience). With "They are here" it is attempted to bring the sonic event to the listener (the performers are put into the listener’s room).

Some of these mixing methods, like creating a completely new event "where anything goes", can result in stereo signals with quite arbitrary relations between the channels.

2.2. Stereo Coding

Stereo coding aims at removing redundancies and irrelevancies from the stereo signal to attain lower bit rates than the sum of the separately coded channels, while maintaining the same quality level and preserving the spatial image. This is done by making use of the correlation between the channels or by exploiting binaural perceptual masking. The known stereo coding methods are:

- Sum Difference coding
- Intensity Stereo Coding
- Stereo Linear Prediction
- Complex Linear Prediction
- VECMA (Very Efficient Coding of Multichannel Audio)
- Parametric Stereo Coding

In the next subsections these techniques are briefly described with their advantages and shortcomings.

2.2.1. Sum-Difference coding

Sum-Difference coding, also known as Mid/Side coding, was first introduced by Johnston [25], as an extension to a transform coder. In this paper, it is proposed not to encode the left and
right signals \((x_1 \text{ and } x_2)\) separately, but its sum \(m\) and its difference \(s\) signal according to

\[
m[n] = x_1[n] + x_2[n] \quad \text{or} \quad m[n] = \frac{1}{2}(x_1[n] + x_2[n])
\]
\[
s[n] = x_1[n] - x_2[n] \quad \text{or} \quad s[n] = \frac{1}{2}(x_1[n] - x_2[n])
\]

The transformation from Left and Right (L/R) to Mid and Side (M/S) (Fig. 2.3) is in the absence of quantization a lossless operation and reversible. Because this is a simple matrix operation, it can be performed in the time or frequency domain.

When left and right signals are similar, applying this procedure results in a mid signal \(m\) with an amplitude approximately twice the size, and a side signal \(s\) with a very low amplitude. Next, the mid signal can be encoded as a mono audio signal. The noisy side signal, however, does not have the same characteristics as an audio signal, and should preferably be coded in another way. Nevertheless because of the low energy of the side signal, not many bits have to be spends on it. The bit rate reduction is about 50\%, however, it is not guaranteed that the side signal always has the lowest energy, e.g. when both signals are opposite in phase then it is the other way around.

A problem can occur when the two stereo channels are completely uncorrelated, we consider the case where one channel virtually silent. The mid and side signals become virtually the same (or opposite in phase). The coding noise introduced by the transform coder is generally masked in the M/S signals but this is not necessarily true for the decoded L/R signals. For the given example, the same amount of coding noise is introduced for the M/S signals. When reconstructing the L/R signals, the noise can become clearly audible in the virtually silent channel. Therefore, it is better not to use the same psycho-acoustical model for the M/S as for the L/R signals.

For conserving the total energy distribution, (2.1) was adapted to normalized sum and difference signals

\[
m[n] = \frac{1}{\sqrt{2}}(x_1[n] + x_2[n]) \quad \text{or} \quad m[n] = \frac{1}{2}(x_1[n] + x_2[n])
\]
\[
s[n] = \frac{1}{\sqrt{2}}(x_1[n] - x_2[n]) \quad \text{or} \quad s[n] = \frac{1}{2}(x_1[n] - x_2[n])
\]

On top of that, the improved Sum-Difference scheme ([26]) was also able to dynamically switch between coding the independent L/R signals or the M/S signals. When the correlation between the stereo signals is high, the M/S signals are encoded, otherwise the L/R signals. In the latest version of M/S coding [33], the psycho-acoustical model is not directly applied to the M/S signals.
2.2. Stereo Coding

2.2.2. Intensity stereo coding

Intensity Stereo Coding (ISC) [23, 24, 27] exploits the irrelevancy of exact location information at higher frequencies. The spatial perception of the human auditory system is frequency dependent: at the lower frequencies, magnitude and phase are dominant spatial cues whereas, at higher frequencies \((f > 2 \text{ kHz})\), spatial perception relies on the analysis of the energy-time envelopes \([6]\). Thus the exact waveforms in the higher frequencies of both channels are not the relevant cues for spatial perception.

Exploiting this was first introduced in \([40]\) as an additional method to increase the quality for a stereophonic subband coder at low bit rates. Initially a rotator (also known as a main axis transform) was used to transform (join) the stereo signals into the so-called intensity (main) signal and error (side) signal. This can be seen as an optimal exploitation of redundancy of sample pairs, but this only leads to marginal gain in bit rate. Therefore, in the higher frequency bands only the intensity signal is coded and transmitted with some overhead in the form of scaling factors. These scaling factors preserve the envelope information in the left and right channel. The method is depicted in Fig. 2.4.

![Figure 2.4: Intensity stereo coding scheme in a high frequency band.](image)

The exact waveform is generally not preserved due to loss of the phase information of the higher frequencies, which can sometimes be perceived as a smaller stereo width. The achieved gain in bit rate is about 20-40%.

In a less complex implementation, the rotator is replaced by the normalized sum signal, this saves computational cost. But there is the possibility for signal cancellation if the signals in the two channels are of opposite sign.

Intensity stereo coding is well suited for pan-potted signals, but it performs poorly with time-delayed signals and for excerpts with fast changes in directional information, e.g. applause. A possible solution is an advanced perceptual algorithm which controls the joint stereo coding.

A similar kind technique for reducing data in a multichannel transmission scheme is introduced by Edler and Fuchs, which is claimed in the patents \([12, 13]\) and also described in \([17]\). It is a post-processing technique which uses cross-channel prediction, and predicts only one channel from the other one (Fig. 2.5). First, the signals are optimally aligned in time by
a varying delay. Next, the control box determines an estimate $\hat{y}[n]$ from $x[n]$. The prediction direction can vary, from the left channel to the right or vice versa. The signals $x[n]$ and $e[n]$ are then transmitted along with the delay $d$ and the prediction coefficients $\alpha_k$. Because of the prediction, the variance of $e[n]$ is lower compared than that of the original signal $y[n]$. The averaged prediction gain achieved with this is between 1.5 and 6.5 dB. The gain is most likely to be less than intensity stereo because of the residual. But the quality on the other hand is better, because less information is discarded.

![Figure 2.5: Cross-channel linear prediction.]

### 2.2.3. Stereo linear prediction

The normal LPC, from appendix B, can be expanded quite straightforward to a full stereo linear prediction. This concept for a stereo audio coder was first introduced by Cambridge and Todd in [10] in 1993. There it was used for a lossless audio coder. The mono predictors (auto-predictors: $P_{11}$, $P_{22}$) are left in place but the so-called cross-predictors ($P_{12}$, $P_{21}$) are added for exploiting the inter-channel correlation (Fig. 2.6).

In this scheme, the next sample from a channel is not only predicted according to its own past but also from the other channel’s past. In particular, the predicted signals according to [10] are given by

\[
\hat{x}_1[n] = \sum_{k=1}^{N_{11}} \alpha_{11,k} x_1[n-k] + \sum_{k=1}^{N_{12}} \alpha_{12,k} x_2[n-k]
\]

\[
\hat{x}_2[n] = \sum_{k=1}^{N_{22}} \alpha_{22,k} x_2[n-k] + \sum_{k=0}^{N_{21}} \alpha_{21,k} x_1[n-k]
\]

where $\hat{x}_1[n]$ and $\hat{x}_2[n]$ are the predicted signals for the left and right channel, the $\alpha$’s are the prediction coefficients and $N$ is the order of the prediction filters. The subscript indicates if it is in the auto or cross predictor branch. The proposed method is not symmetrical (cf eq. 2.3a,b)
because $x_2$ is also estimated from the current sample of $x_1$ (i.e., a direct feed-through). The
residuals are defined by

$$
e_1[n] = x_1[n] - \hat{x}_1[n]$$

$$
e_2[n] = x_2[n] - \hat{x}_2[n]$$

Thus, if both signals are identical, this results in $\alpha_{21,0}$ being equal to 1, and the residual $e_2$ will
be zero.

The parameters $\alpha$, for auto and cross predictor, are optimized simultaneously but separated
per channel, this because of the non-symmetry.

Experiments with this stereo predictor [10, 18] were conducted to compare the achieved gain
with two mono predictors. Fixed orders for the mono and stereo predictors were used, but
keeping the total order the same, i.e. mono order: $2 \times 12$; stereo order: $2 \times (8_{auto} + 4_{cross})$. The
averaged additional gain achieved, for lossless backward stereo prediction, is 0.3 bits/sample,
but the gain is very signal dependent e.g. a mono-like excerpt can achieve a higher gain because
of the direct feed-through. Thus, it was concluded that the reduction in bit rate, in comparison
with conventional linear prediction techniques, is not significant.

Later in [31] an adaptive method for optimizing was proposed which first determines the auto
prediction order by optimize the mono predictors by normal Levinson-Durbin, subsequently the
cross prediction order is increased until the bit rate reaches a minimum. Finally, the algorithm
verifies if the auto prediction order can be decreased, giving an additional gain. The stereo
scheme is also generalized to multichannel prediction, with mutual prediction of any number of
channels.

More recently, a new generalized stereo decorrelation algorithm was proposed in [20], which
calculates the optimal parameters, but still separating left and right.
Finally, in [19] block Levinson [36] is proposed for the non-direct feed-through branch, under the restriction of equal orders $N_{11} = N_{12}$. In here, also forward adaptation is compared with backward adaptation. Backward adaptation is used, such that no parameters $\alpha$ have to be transmitted. A proper method for quantizing these coefficients is not known.

### 2.2.4. Complex linear prediction

In [21], an extension is proposed from Warped Linear Predictor (WLP), a monophonic perceptual audio codec, to a warped linear predictor for Complex valued signals (CWLP). Combining the left ($x_1$) and right ($x_2$) channel into one complex signal can be done in different ways. Three methods were proposed:

\[
\begin{align*}
     c_1[n] &= x_1[n] + j x_2[n] \quad (2.5) \\
     c_2[n] &= (x_1[n] + x_2[n]) + \frac{1}{2} j (x_1[n] - x_2[n]) \quad (2.6) \\
     c_3[n] &= (x_1[n] - H\{x_1[n]\}) + \frac{1}{2} j (x_2[n] - H\{x_2[n]\}) \quad (2.7)
\end{align*}
\]

where $H$ denotes the Hilbert transformation.

A constraint which is imposed by combining the left and right signal into a complex signal is a limitation in the available degrees of freedom for the predictor. Roughly, one could say that for a 2-channel signal, we have the following degrees of freedom:

- $r_{11} \rightarrow$ symmetric, $N$ degrees;
- $r_{22} \rightarrow$ symmetric, $N$ degrees;
- $r_{12} \rightarrow$ arbitrary, $2N$ degrees,

with $r_{ij}$ the correlation function. For a complex signal, however, we have

- $Re\{r_{cc}\} \rightarrow$ symmetric, $N$ degrees;
- $Im\{r_{cc}\} \rightarrow$ anti-symmetric, $N$ degrees,

with $r_{cc}$ the correlation function of the complex signal.

This means that making a complex signal out of two real signals and then doing an optimization on basis of the autocorrelation function must be suboptimal because already half of the information of the relation between the two signals is lost due to merging these into one complex signal.

### 2.2.5. Rotation (VECMA)

Very Efficient Coding of Multichannel Audio (VECMA) [34] is another method which aims at efficiently coding the stereo audio. It uses rotation (also known as Principle Component Analysis, PCA), not like in intensity stereo coding (see subsection 2.2.2) on the complete frequency band but on subbands. VECMA aims at separating sound sources.

The general scheme of the VECMA coder is drawn in Fig. 2.7, PCA determines the angle at which the source is located, and the stereo plane is rotated over this angle, resulting in two new signals: a dominant signal $m$ containing the sound sources, and a residue signal $s$ containing the accompanying reverberation. The residue signal is considered redundant, and is discarded. Only the energy levels, which can be coded more efficient, are transmitted this along with the angles. The discarded signal is synthetically generated at the decoder side by reverberating the dominant signal, and subsequently scaling to the correct energy levels.
2.2. Stereo Coding

VECMA is able to reconstruct the positions of the sound sources close to their original position in the stereo image, but the simplifying assumption that the residue only contains the reverberation results in an overall narrower stereo image.

2.2.6. Parametric stereo coding

The most recent developed stereo coding method is parametric stereo coding, which aims at separating the audio content and the auditory spatial information. The perceptually relevant spatial sound cues are extracted by analyzing the two channels of the stereo audio and, subsequently, the stereo channels are mixed down to one channel. An additional advantage for mixing to a mono signal is that a normal audio coder can be used for coding this signal. As such, it can be seen as an enhancement for existing mono audio codecs, providing stereo audio at low to very low bit rates.

Binaural Cue Coding (BCC) is such a parametric coding technique and was introduced by Faller and Baumgarte [3, 15]. BCC is based on the assumption that in every critical band the dominant source defines the spatial perception. A compact parametric representation of the auditory spatial information is made by extracting Inter-Channel Level Differences (ICLDs) and Inter-Channel Time Differences (ICTDs) which are the signal equivalents of ILDs and ITDs, respectively. From the stereo signal, a mono signal is made by adding both channels, and this is coded with a mono audio coder.

The generic BCC scheme for coding a stereo signal is shown in Fig. 2.8. The BCC analyzer extracts per critical band, ICLD and ICTD. These are transmitted along with the encoded mono signal. At the decoder the spatial sound image is reconstructed by applying ICLDs and ICTDs but, because of down-mix, the signal reconstruction is not perfect.

The BCC schemes can capture the location of various sound sources quite effectively, but the reconstructed spatial image lacks spatial ambience, such as reverberation and spatial diffuseness. Therefore, an extension to this spatial coding scheme has been proposed, namely Optimum Coding of Stereo (OCS) which is developed within Philips by J. Breebaart et.al. [7]. Herein a third sound-field parameter is added, i.e., inter-channel coherence, which is able to capture these ambience properties. In the more recent BCC scheme [4, 16], a coherence parameter is
The OCS scheme is generally the same as BCC (Fig. 2.8), the parameters are extracted from the stereo signal. The power ratio of the corresponding subband $b$ is defined as

$$\text{HRD}[b] = 10 \log_{10} \frac{\sum_k X_1[k]X_1^*[k]}{\sum_k X_2[k]X_2^*[k]}$$  \hspace{1cm} (2.8)$$

with $X$ the DFT of signal $x$ and where the summations extend over all $k$'s within the subband $b$. The inter-channel phase difference is obtained as follows

$$\text{IPD}[b] = \angle \left( \frac{\sum_k X_1[k]X_2^*[k]}{\sum_k X_2[k]X_1^*[k]} \right) .$$  \hspace{1cm} (2.9)$$

The inter-channel coherence is defined as

$$\text{IC}[b] = \frac{\left| \sum_k \tilde{X}_1[k]X_2^*[k] \right|}{\sqrt{\left( \sum_k \tilde{X}_1[k]X_1^*[k] \right)^2 \left( \sum_k \tilde{X}_2[k]X_2^*[k] \right)^2}}$$  \hspace{1cm} (2.10)$$

with $\tilde{X}_1$ and $\tilde{X}_2$ the DFT of the signals after phase alignment according to the estimated IPD.

Next to the additional IC parameter, another difference with the BCC scheme is that the down-mix is more refined. Furthermore, an Overall Phase Difference (OPD) is transmitted which enables reconstructing with the proper alignment.

Notwithstanding this success of OCS for the low bit rates, it appears to be difficult to attain transparency at high bit rates when using this technique. This is presumably due to the imposed model.
2.2. Stereo Coding

Confidential
3. Stereo LPC

As mentioned in section 2.2.3, Stereo Linear Prediction was introduced in [10] as a lossless audio compression method. Here a slightly different stereo coding scheme is proposed. It uses a 2-channel or stereo linear prediction in combination with a single rotator (Fig. 3.1). The stereo LP still consists of two auto-predictors ($P_{11}$ and $P_{22}$ in Fig. 3.1) and two cross-predictors ($P_{12}$ and $P_{21}$), for exploiting, respectively, intra and inter channel correlations. This scheme differs in two aspects from that in [10]. First, the prediction scheme is *symmetrical* from left to right, and vice versa, and second, the addition of a *rotator*.

The complete scheme of the 2-channel codec exists of an analysis (Fig. 3.1) and a synthesis (Fig. 3.2) system, both of which will be explained in the next section. Moreover, the predictor filters $P_{ij}$ are discussed, and finally a method for calculating the optimal parameter is described along with some problem areas and possible solutions.

![Figure 3.1.: Analysis system: Stereo Linear Prediction in combination with a single rotator (R).](image-url)
3.1. Two-channel LP and rotation

In this section, we describe the stereo LP analysis and synthesis and rotation systems.

Analysis system

The analysis system, or encoder, (Fig. 3.1) exist of two basic blocks, namely a stereo linear predictor and a rotator (R).

The Stereo Linear Predictor removes redundancies which are present in the input signals \(x_1\) and \(x_2\), by making estimate signals \(\hat{x}_1\) and \(\hat{x}_2\) from the previous (filtered) samples of \(x_1\) and \(x_2\). The estimation for the left and right channel is abstractly given by

\[
\hat{x}_1[n] = \sum_{k=1}^{N_a} \alpha_{11,k} y_{1,k}[n] + \sum_{k=1}^{N_c} \alpha_{12,k} y_{2,k}[n]
\]

(3.1a)

\[
\hat{x}_2[n] = \sum_{k=1}^{N_a} \alpha_{21,k} y_{1,k}[n] + \sum_{k=1}^{N_c} \alpha_{22,k} y_{2,k}[n]
\]

(3.1b)

with prediction coefficients \(\alpha_{ij,k} (i, j = 1, 2)\), auto prediction order \(N_a\), cross order \(N_c\) and the weighted input signals \(y_{i,k} = f_k * x_i\), where \(f_k\) are the impulse responses of causal stable filters with no direct feed-through. The summations with coefficients \(\alpha_{11}\) and \(\alpha_{22}\) indicate the auto-predictors, and the summations with coefficients \(\alpha_{12}\) and \(\alpha_{21}\) the cross-predictors. So unlike the scheme in [10], no direct feed-through is used.

The residual or prediction errors \((e_1\) and \(e_2\)) originate from subtracting the estimated signals from the original ones:

\[
e_1[n] = x_1[n] - \hat{x}_1[n] = x_1[n] - \sum_{k=1}^{N_a} \alpha_{11,k} y_{1,k}[n] - \sum_{k=1}^{N_c} \alpha_{12,k} y_{2,k}[n],
\]

(3.2a)

\[
e_2[n] = x_2[n] - \hat{x}_2[n] = x_2[n] - \sum_{k=1}^{N_a} \alpha_{21,k} y_{1,k}[n] - \sum_{k=1}^{N_c} \alpha_{22,k} y_{1,k}[n].
\]

(3.2b)

In vector notation, and with the condition of equal orders i.e. \(N_a = N_c = N\), this can be written as

\[
\mathbf{e}[n] = \mathbf{x}[n] - \hat{\mathbf{x}}[n] = x[n] - \sum_{k=1}^{N} \mathbf{A}_k \mathbf{y}[n] = \left[ x_1[n] \ x_2[n] \right] - \sum_{k=1}^{N} \left[ \begin{array}{cc} \alpha_{11,k} & \alpha_{12,k} \\ \alpha_{21,k} & \alpha_{22,k} \end{array} \right] \left[ \begin{array}{c} y_{1,k}[n] \\ y_{2,k}[n] \end{array} \right]
\]

(3.3)

Transformed to frequency domain this will result in

\[
E(z) = H(z) \cdot X(z) = \left[ \begin{array}{cc} 1 - P_{11}(z) & -P_{12}(z) \\ -P_{21}(z) & 1 - P_{22}(z) \end{array} \right] \left[ \begin{array}{c} X_1(z) \\ X_2(z) \end{array} \right]
\]

(3.4)
with $X(z)$ and $E(z)$ the $z$-transforms of the input and residual signals, and with the total transfer matrix $H(z)$ which consists of the individual predictor transfers $P_{ij}(z)$ defined by

$$P_{ij}(z) = \sum_{k=1}^{N} \alpha_{ij,k} F_k(z). \tag{3.5}$$

Assuming a sufficiently high order $N$, the analysis results in residual signals $(e)$ which have a lower variance than the original $(x)$. Since all redundancies are removed, the residual can be coded more efficiently. The stereo linear predictor also ensures that the spectra of the residual signals are flattened and that the cross-correlation between these signals have been minimized, except for lag 0.

For removing the correlation between the sample pairs, a single rotator, also known as an axis transform, is used on the residual signal. A single rotator should be sufficient now because other relations have already been removed by the linear predictor stage. Rotation with angle $\varphi$, applied on the residual signals, is a simple matrix multiplication described by

$$\begin{bmatrix} m \\ s \end{bmatrix} = R(\varphi) \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} e_1 \cdot \cos(\varphi) - e_2 \cdot \sin(\varphi) \\ e_2 \cdot \sin(\varphi) + e_1 \cdot \cos(\varphi) \end{bmatrix} \tag{3.6}$$

The rotator constructs a main $m$ and a side $s$ signal which are uncorrelated for lag 0. The rotation angles are transmitted as side information to the synthesis system. The total analysis scheme therefore results in two uncorrelated, approximately flat signals.

**Synthesis system**

The synthesis system, or decoder (Fig. 3.2) performs the opposite processing of the analysis system thereby reconstructing the original signals. First, the inverse rotation (IR) is applied undoing the rotation in the analysis system. It is obvious that inverse rotation is a rotation over the negative angle ($R(\varphi)^{-1} = R(-\varphi)$).

The LP synthesis filter uses the same predictors ($P_{ij}$) as analysis system but now in a feedback loop instead of a feed-forward loop. Since the filter transfers $P_{ij}(z)$ have no direct feed-through, this system is implementable in the manner as shown in Fig. 3.2.

The transfer of the LP synthesis system is described by $\{H(z)^{-1}\}$. For the two-channel case this results in

$$\{H(z)^{-1}\} = \frac{1}{\det(H(z))} \begin{bmatrix} 1 - P_{22}(z) & P_{12}(z) \\ P_{21}(z) & 1 - P_{11}(z) \end{bmatrix} \tag{3.7}$$

with the determinant of $H(z)$ as: $\det(H(z)) = (1 - P_{11}(z))(1 - P_{22}(z)) - P_{21}(z)P_{12}(z)$. From this we observe that the stability of the LP synthesis filter ($\{H(z)^{-1}\}$) is guaranteed if $\frac{1}{\det(H(z))}$ is a stable filter. So, all the poles of the synthesis system are determined by the determinant of $H(z)$ and, to ensure a stable synthesis filter, all poles have to lie within the unit circle. The stability issue is further discussed in section 3.4.
3.2. Prediction filters

In this section, the prediction filter structure will be considered in more detail. We start from Fig. 3.1, which is somewhat more general than equations (3.1). The partial prediction filters \( P_{ij} \) (Fig. 3.1 and 3.2) can be written as:

\[
P_{ij}(z) = \sum_{k=1}^{N} \alpha_{ij,k} F_k(z)
\]

i.e. a pre-filter \( F_0^{(ij)} \) (inserted to prevent direct feed-throughs) multiplied by a weighted sum of filters arbitrary causal and stable filters with transfers \( F_m^{(ij)} \) \((m = 0..N)\). In view of symmetry, \( F_m^{(ij)} = F_m^{(ji)} \) is taken. Furthermore, in order to reduce complexity \( F_m^{(ij)} = F_m^{(ii)} \) is taken. Hence, superscripts \((i,j)\) can be dropped, and by defining \( F_k(z) = F_0(z) F_k^{(ii)}(z) \) to keep consistent with (3.5), this results in:

\[
P_{ij}(z) = \sum_{k=1}^{N} \alpha_{ij,k} F_k(z)
\]

This symmetrical form is the most efficient filter structure, because now the same intermediate signals \( y_{ik,k} \) can be used for the partial predications \( p_{ij} \) (Fig. 3.3) of left and right signal. This scheme is also the most effective when for auto and cross predictors equal prediction order is used.

A similar realization as in Fig. 3.3 can be made for the synthesis filter, but now using the filters \( F_0 \) and \( F_k \) in a feedback loop.
Figure 3.3.: Symmetrical filter implementation of a two-channel analysis predictor.
The simplest choice for $F_k(z)$ would be to use a Tapped-Delay-Line (TDL) as filter structure, i.e., $F_k(z) = z^{-k}$. In this case (3.4) can be rewritten, using (3.9) and the TDL transfer, to a compact notation

$$
\varepsilon[n] = \sum_{k=0}^{N} A'_k z^n \quad \text{or as transfer } \quad H(z) = \sum_{k=0}^{N} A'_k z^{-k}
$$

(3.10)

with $A'_0 = I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $A'_i = -A_i$ for $i = 1..N$, still assuming equal orders ($N_a = N_c = N$).

From [8, 22], we know that for audio signals it is advantageous to take a warped filter structure, according to a psycho-acoustical frequency scale (e.g., the Bark Scale) by using Laguerre filters. In the case of Laguerre filters, the filters $F_k$ are given by

$$
F_k(z) = z^{-1} \frac{\sqrt{1 - \lambda^2}}{1 - z^{-1} \lambda} \left( \frac{-\lambda + z^{-1}}{1 - \lambda z^{-1}} \right)^{k-1}
$$

(3.11)

with Laguerre coefficient $|\lambda| < 1$. A warping factor $\lambda = 0.7564$ (for sample frequency $F_s = 44.1$ kHz) gives a good correspondence to a Bark scale [39]. Performing LP on this warped scale places more emphasis on accurate modeling of the spectral envelope of the input signal at the lower frequencies. $\lambda = 0$ results in a TDL, like discussed above.

For the Laguerre system, the resulting transfer is given by

$$
H(z) = I - z^{-1} \frac{\sqrt{1 - \lambda^2}}{1 - z^{-1} \lambda} \sum_{k=1}^{N} A_k \left( \frac{-\lambda + z^{-1}}{1 - \lambda z^{-1}} \right)^{k-1}
$$

(3.12)

Conform [8], this expression can be rewritten to an all-pass line with the following notation

$$
H(z) = \sum_{k=0}^{N} B_k \left( \frac{-\lambda + z^{-1}}{1 - \lambda z^{-1}} \right)^{k}
$$

(3.13)

with a linear mapping from $A_k$ to $B_k$.

### 3.3. Calculation of the optimal parameters

The outputs of the analysis system, as depicted in Fig. 3.1, are the main $m$ and side $s$ signal. Optimizing the parameters, i.e., the prediction coefficients and rotation angle, can be defined as an optimization problem which produces a maximum of a squared sum of the main signal (max $\sum \{ |m[n]|^2 \}$), which automatically induces a minimum for the squared sum of the side signal (min $\sum \{ |s[n]|^2 \}$).

It can be proved (see appendix E) that joint optimization of prediction coefficients $\alpha$'s and rotation angle $\varphi$ is equal to sequential optimization of $\alpha$'s based on minimization of the squared sum of the residual signals (min $\sum \{ |e_1[n]|^2 \}$ and min $\sum \{ |e_2[n]|^2 \}$), followed by PCA on the residuals $e_1$ and $e_2$. The calculation of the optimal prediction coefficients is now considered in more detail.
Optimal stereo-LPC coefficients

For calculation of the optimal stereo LP coefficients, we can consider $\min \{ \sum_n |e_1[n]|^2 \}$ where $e_1^2$ is determined by coefficients $\alpha_{11,k}$ for $k = 1..N_a$, and $\alpha_{12,k}$, for $k = 1..N_c$. For convenience we introduce the following vectors:

$$
\bar{e}_i = [e_i[0], e_i[1], \cdots, e_i[L]]^t \text{ for } i = 1, 2,
$$

$$
\bar{x}_i = [x_i[0], x_i[1], \cdots, x_i[L]]^t \text{ for } i = 1, 2,
$$

$$
\bar{y}_{i,k} = [y_{i,k}[0], y_{i,k}[1], \cdots, y_{i,k}[L]]^t \text{ for } i = 1, 2 \text{ and } k = 1, \cdots, \max(N_a, N_c)
$$

with data length $L$, and

$$
\alpha_{ii} = [\alpha_{ii,1}, \alpha_{ii,2}, \cdots, \alpha_{ii,N_a}]^t \text{ for } i = 1, 2 \text{ (3.17)}
$$

$$
\alpha_{ij} = [\alpha_{ij,1}, \alpha_{ij,2}, \cdots, \alpha_{ij,N_c}]^t \text{ for } i, j = 1, 2 \text{ and } i \neq j \text{ (3.18)}
$$

Furthermore, we introduce the matrices $Y_i (i, 1, 2)$ by

$$
Y_1 = [\bar{y}_{1,1}, \bar{y}_{1,2}, \cdots, \bar{y}_{1,N_a}, \bar{y}_{2,1}, \bar{y}_{2,2}, \cdots, \bar{y}_{2,N_c}] \text{ (3.19)}
$$

$$
Y_2 = [\bar{y}_{1,1}, \bar{y}_{1,2}, \cdots, \bar{y}_{1,N_a}, \bar{y}_{2,1}, \bar{y}_{2,2}, \cdots, \bar{y}_{2,N_c}] \text{ (3.20)}
$$

From the previous definitions, it follows that

$$
\bar{e}_1 = \bar{x}_1 - Y_1 \begin{bmatrix} \alpha_{11} \\ \alpha_{12} \end{bmatrix} \text{ (3.21)}
$$

where $[\alpha_{11} \alpha_{12}]$ is a stacked vector of auto and cross prediction coefficients. Thus, the optimal $\alpha$, is given by the normal equations

$$
[\begin{bmatrix} \alpha_{11} \\ \alpha_{12} \end{bmatrix}] = \{Y_1^t Y_1\}^{-1} \{Y_1^t \bar{x}_1\} \text{ (3.22)}
$$

From minimizing $\sum_n |e_1^2[n]|$, a similar normal equation can be obtained for the remaining coefficients:

$$
[\begin{bmatrix} \alpha_{21} \\ \alpha_{22} \end{bmatrix}] = \{Y_2^t Y_2\}^{-1} \{Y_2^t \bar{x}_2\} \text{ (3.23)}
$$

Calculation of the optimal $\alpha$'s can be done by directly solving these two sets of normal equations ((3.22) and (3.23)), but this is not always the most efficient way. For the condition of equal order $N_a = N_c = N$, we have that $Y_1 = Y_2 = Y$ resulting in the following set of equations

$$
[\begin{bmatrix} \alpha_{11} & \alpha_{21} \\ \alpha_{12} & \alpha_{22} \end{bmatrix}] = \{Y^t Y\}^{-1} \{Y^t [\bar{x}_1, \bar{x}_2]\} \text{ (3.24)}
$$
3.3. Calculation of the optimal parameters

For equal order, the square matrix $Y^t Y$ consists of 4 Toeplitz matrices:

\[
Y^t Y = \begin{bmatrix}
R_{11,0} & R_{11,-1} & \cdots & R_{11, -(N-1)} & R_{21,0} & R_{21,-1} & \cdots & R_{21, -(N-1)} \\
R_{11,1} & R_{11,0} & \cdots & R_{11, -(N-2)} & R_{21,1} & R_{21,0} & \cdots & R_{21, -(N-2)} \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
R_{11,N-1} & R_{11,N-2} & \cdots & R_{11,0} & R_{21,1} & R_{21,0} & \cdots & R_{21,0} \\
R_{12,0} & R_{12,-1} & \cdots & R_{12, -(N-1)} & R_{22,0} & R_{22,-1} & \cdots & R_{22, -(N-1)} \\
R_{12,1} & R_{12,0} & \cdots & R_{12, -(N-2)} & R_{22,1} & R_{22,0} & \cdots & R_{22, -(N-2)} \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
R_{12,N-1} & R_{12,N-2} & \cdots & R_{12,0} & R_{22,1} & R_{22,0} & \cdots & R_{22,0}
\end{bmatrix}
\]

(3.25)

with

\[
R_{ij,k} = \sum_{k=-(N-1)}^{N-1} y_i[n]y_j[n+k]
\]

(3.26)

If we now define a $2 \times 2$ block matrix (is not necessarily a Toeplitz), as

\[
R_k = \begin{bmatrix}
R_{11,k} & R_{12,k} \\
R_{21,k} & R_{22,k}
\end{bmatrix}
\]

(3.27)

the matrix $Y^t Y$ can be reorganized into a block-Toeplitz structure according to

\[
\begin{bmatrix}
R_0 & R_{-1} & R_{-2} & \cdots & R_{-(N-1)} \\
R_1 & R_0 & R_{-1} & \cdots & R_{-(N-2)} \\
R_2 & R_1 & R_0 & \cdots & R_{-(N-3)} \\
\vdots & \ddots & \ddots & \ddots & \ddots \\
R_{N-1} & R_{N-2} & R_{N-3} & \cdots & R_0
\end{bmatrix}
\]

(3.28)

Symmetry of block-Toeplitz matrix is not guaranteed because $R_k = R_{-k}^t$ but generally not $R_k \neq R_{-k}$.

The normal equation (3.24) can thus be written as

\[
\begin{bmatrix}
A_1 & \cdots & A_N
\end{bmatrix}
\begin{bmatrix}
y_1 \\
y_2 \\
y_3 \\
\vdots
\end{bmatrix}
= \begin{bmatrix}
y_0 \\
R_1 & R_0 & R_{-1} & \cdots & R_{-N+1} \\
R_1 & R_0 & R_{-1} & \cdots & R_{-N+2} \\
R_1 & R_2 & R_0 & \cdots & R_{-N+3} \\
\vdots & \ddots & \ddots & \ddots & \ddots \\
R_{N-1} & R_{N-2} & R_{N-3} & \cdots & R_{0}
\end{bmatrix}
\begin{bmatrix}
A_1 \\
A_2 \\
A_3 \\
\vdots
\end{bmatrix}
\]

(3.29)

with equal-sized square matrices $R_k$, $A_k$ and $y_k$, and where the prediction coefficients $A_k$ are the unknowns to be solved. An iterative Levinson [37] type of algorithm can be used, namely the block-Levinson algorithm by Wiggins and Robinson (also known as Levinson-(Whittle)-Wiggins-Robinson or LWR algorithm [41, 36]) which is an efficient way for solving this.

A difference with the normal (one-channel) Levinson algorithm is that, in the multichannel-case, matrix operations appear and special care has to be taken since matrix multiplications
do not commute. An additional advantage of this iterative structure is that we can build in stop conditions to prevent unstable synthesis filters or to stop when the gain is not sufficiently improving. Using such stop criteria results in an adaptive order.

The block-Levinson algorithm derivation follows closely that in Numerical Recipes [37], but is here adapted for blocks matrices.

The algorithm’s iterative order (i.e., iteration step) is denoted by $M$. The value of $M$ is increased until $N$ is reached (or another stop condition is met).

The fundamental equations to be solved read

$$
\sum_{j=1}^{M} R_{i-j} A_{j}^{(M)} = y_{i} \quad \text{for} \quad i = 1, \ldots, M.
$$

(3.30)

Initializing the block-Levinson for first order ($M = 1$) is done by

$$
A_{1}^{(1)} = R_{0}^{-1} y_{1}, \quad \ \ \ \ \ (3.31a)
$$

$$
G_{1}^{(1)} = R_{0}^{-1} R_{-1}, \quad \ \ \ \ \ (3.31b)
$$

$$
H_{1}^{(1)} = R_{0}^{-1} R_{1}, \quad \ \ \ \ \ (3.31c)
$$

where $G_{j}^{(M)}$ are the backward reflection matrices, and $H_{j}^{(M)}$ are the forward reflection matrices at iteration $M$ for $j = 1, \ldots, M$. Next, these reflection matrices are updated for iteration $M + 1$ by

$$
G_{M+1}^{(M+1)} = \left[ \sum_{j=1}^{M} R_{j-(M+1)} H_{j}^{(M)} \right]^{-1} \left[ \sum_{j=1}^{M} R_{j-(M+1)} G_{j}^{(M)} - R_{-(M+1)} \right],
$$

(3.32a)

$$
H_{M+1}^{(M+1)} = \left[ \sum_{j=1}^{M} R_{M+1-j} G_{M}^{(M)} - R_{0} \right]^{-1} \left[ \sum_{j=1}^{M} R_{M+1-j} H_{j}^{(M)} - R_{M+1} \right],
$$

(3.32b)

Subsequently, the reflection matrices for the next iteration $M + 1$ can be determined (this only involves the previous reflection matrices):

$$
H_{j}^{(M+1)} = H_{j}^{(M)} - G_{M+1-j}^{(M+1)} H_{j}^{(M+1)}
$$

(3.33a)

$$
G_{j}^{(M+1)} = G_{j}^{(M)} - H_{M+1-j}^{(M)} G_{M}^{(M+1)}
$$

(3.33b)

Finally, the prediction coefficients can be updated through:

$$
A_{M+1}^{(M+1)} = \left[ \sum_{j=1}^{M} R_{M+1-j} G_{M+1-j}^{(M)} - R_{0} \right]^{-1} \left[ \sum_{j=1}^{M} R_{M+1-j} A_{j}^{(M)} - y_{M+1} \right],
$$

(3.34a)

$$
A_{j}^{(M+1)} = A_{j}^{(M)} - G_{M+1-j}^{(M)} A_{M+1}^{(M+1)}, \quad (j = 1, \ldots, M).
$$

(3.34b)
Optimal rotation angle

The optimal angle for the rotator can be determined as described in appendix C. In summary, the algorithm is as follows:

\[
R_{11} = \vec{e}_1^T \cdot \vec{e}_1 \tag{3.35}
\]

\[
R_{22} = \vec{e}_2^T \cdot \vec{e}_2 \tag{3.36}
\]

\[
R_{12} = \vec{e}_1^T \cdot \vec{e}_2 = \vec{e}_2^T \cdot \vec{e}_1 \tag{3.37}
\]

\[
\mu_{mean} = \frac{1}{2} (R_{11} + R_{22}) \tag{3.38}
\]

\[
c = \frac{1}{2} (R_{22} - R_{11}) + jR_{11}. \tag{3.39}
\]

The angle of \( c \) is a measure for the optimal rotation \( \phi \), in particular

\[
\phi = \frac{1}{2} \angle c + k\pi \text{ with } k = \cdots, -1, 0, 1, 2, \cdots. \tag{3.40}
\]

For later use, we define the modulation depth \( m \) by

\[
m = \frac{|c|}{\mu_{mean}} \in [0, 1]. \tag{3.41}
\]

3.4. Practical problems and solutions

From the above definitions of calculation of optimal parameters, it can be shown that some input signals can cause problems, numerical or otherwise. This is also confirmed through the preformed experiments.

For instance, consider \( R_0^{-1} \) in the initialization of the block-Levinson algorithm (3.31a), for mono-like signals, where \( x_1[n] \approx x_2[n] \). It is not unambiguously defined how to predict because auto and cross correlation are nearly equal for mono-like signals, resulting in an ill-condition inversion of \( R_0 \). In [10] this does not occur because of the direct feed-through, but the optimization in here is separated for the left and right signals. An experiment was conducted, trying to circumvent this, by applying a pre-rotator (before the prediction). Applying rotation on a mono-like signal results in a dominant main signal and a side signal which is almost zero. Next, this main and side signal are fed into the stereo LP. However, pre-rotation does not change the condition number of \( R_0 \); it remains ill-conditioned. Another solution needs to be found.

A zero signal in a channel can also occur without a pre-rotator, e.g., a short silence in one channel, which will cause the same ill-conditioned problem.

The last input signal, discussed in this section, which can cause trouble is a uncorrelated stereo signal with equal power. Inversions of matrices in the Block-Levinson are all well-conditioned (only auto prediction). But the problem can occur in the rotator optimisation. For equal-power, uncorrelated input signals, every rotation angle is optimal.

In the next subsections, first, the stability is discussed in more detail, and subsequently possible solutions for the above mentioned problem areas are proposed.
3.4.1. Stability of the 2-channel synthesis filter

First of all, stability of the whole system is determined by the stability of the stereo synthesis LP filter, and the stability of this synthesis filter, as mentioned before, is determined by the roots of \( \det(H(z)) \). This means that the synthesis system is stable if \( \det(H(z)) \neq 0 \) for \(|z| > 1\).

According to Whittle [42], a stable solution is guaranteed when the block-Levinson algorithm is used. However, this algorithm only applies for equal orders \( (N_a = N_c) \) and gives no information concerning stability for unequal orders \( N_a \neq N_c \) (and \( N_c \neq 0 \)).

Experiments have been conducted, optimizing the system by solving the normal equations (3.22) and (3.23) for different orders. After analysis the prediction coefficients were examined, more specific the roots of the determinant of \( H(z) \). Plotted in Fig. 3.4 is the maximum magnitude of these roots per frame. Fig. 3.4a depicts the result for order \( N_a = N_c = 10 \), and the horizontal dashed line indicates the maximum for which stability of the synthesis filter is guaranteed. The experiment confirms that, for equal order, the synthesis filter is stable for all frames. In Fig. 3.4b the results are given for the case that \( N_a = 10 \) and \( N_c = 3 \). It is clear that not in all frames a stable inverse system occurs.

For experimentally finding the stable region, tests were performed analyzing different excerpts with varying cross-prediction orders and a fixed auto prediction order [5]. Result are shown in Fig. 3.5 for \( N_a = 10 \) and \( N_c = 1..20 \). For cross-orders of \( N_c = 7 \) until \( N_c = 11 \), a stable inverse system was found. However, such results are excerpt dependent, but all have the same trend, i.e., stable around the point of equal order and unstable regions before and after this point, where the width of stable region differs for various excerpts. We infer from this that stability of the synthesis filter can only be unconditionally guaranteed for \( N_c = 0 \) or \( N_c = N_a \).

![Figure 3.4.](image)

Figure 3.4.: Maximum of the pole radius of \( \{H(z)\}^{-1} \) (a) for \( N_a = N_c = 10 \) and (b) for \( N_a = 10 \) and \( N_c = 3 \). Used frame size is 1000 samples.
3.4. Practical problems and solutions

3.4.2. Regularization of the Stereo LP optimization problem

Some input signals can lead to ill-conditioned inversions of matrices, i.e., mono-like signals and a zero signal. A solution can be applying a kind of regularization.

Before analyzing, uncorrelated white noise can be 'added' to cancel the strong mono-like character and, thus, improve the conditions of the matrices before inverting them. These white-noise corrections can be achieved through adding noise in a relative ($\epsilon_{rel}$) or absolute ($\epsilon_{abs}$) manner in the normal equations. For the normal equations ((3.22) and (3.23)), this results in

$$Y_i^tY_i \rightarrow Y_i^tY_i + \text{diag}(Y_i^tY_i) \cdot \epsilon_{rel} + I \cdot \epsilon_{abs}$$

(3.42)

where $\text{diag}(Y_i^tY_i)\epsilon_{rel}$ ($0 < \epsilon_{rel} < 1$) is introduced for adding 'white' noise proportionally to the existing auto-correlations ($\text{diag}(Y_i^tY_i)$ indicating a matrix with zeros outside the diagonal and on it the diagonal of $Y_i^tY_i$), and $I \cdot \epsilon_{abs}$ is introduced for adding an absolute amount of noise to both channels.

For the block-Levinson algorithm (3.31a), regularization can be done by checking for singularities before inversion, and subsequently biasing in a similar manner as described above.

3.4.3. Robustness of the stereo LP system

As seen in Fig. 3.4a, even with equal orders, the maximum magnitude of the poles for most of the frames is very close to unity. So in order to make the analysis more robust Spectral Smoothing Technique (SST) is applied. This a known technique from speech coding [29], and is also known as bandwidth extension, bandwidth expansion or bandwidth widening.

Consider the all-pole synthesis filter, which arises from the standard one-channel LP analysis:

![Figure 3.5: Maximum of the pole radius for fixed $N_a = 10$ and varying cross orders $N_c = 1..20$.](image-url)
Applying SST on the predictor coefficients amounts to replacing $\alpha_k$ by $\alpha'_k = \gamma^k \alpha_k$ giving a transfer $H'(z)$ according to

$$H'(z) = \frac{1}{A'(z)} = \frac{1}{1 - \sum_{k=1}^{N} \gamma^k \alpha_k z^{-k}} = \frac{1}{1 - \sum_{k=1}^{N} \alpha_k \left(\frac{\xi}{\gamma}\right)^{-k}} = \frac{1}{A\left(\frac{\xi}{\gamma}\right)} = H\left(\frac{z}{\gamma}\right)$$

(3.44)

where $\gamma$ is the smoothing factor ($0 < \gamma \leq 1$). All the poles of the new filter $H'(z)$ have shifted with a factor $\gamma$ towards the origin (see Fig. 3.6), improving the stability. This at the expense of some of the whitening properties and the overall gain that can be achieved by the analysis filter.

![Figure 3.6: Spectral smoothing shifts the poles towards zero.](image)

This technique can also be extended for the two-channel case. Now consider the transfer for the two-channel case

$$H(z) = \begin{bmatrix} H_{11}(z) & H_{12}(z) \\ H_{21}(z) & H_{22}(z) \end{bmatrix} \text{ with } H_{ij}(z) = \frac{1}{A_{ij}(z)} = \frac{1}{1 - \sum_{k=1}^{N} \alpha_{ij,k} z^{-k}}$$

(3.45)

Applying SST to the individual transfer functions, all with the same $\gamma$ yields

$$H'(z) = \begin{bmatrix} H'_{11}(z) & H'_{12}(z) \\ H'_{21}(z) & H'_{22}(z) \end{bmatrix} = \begin{bmatrix} H_{11}(\frac{\xi}{\gamma}) & H_{12}(\frac{\xi}{\gamma}) \\ H_{21}(\frac{\xi}{\gamma}) & H_{22}(\frac{\xi}{\gamma}) \end{bmatrix}$$

(3.46)

Consider now the determinant of $H'(z)$:

$$\det(H'(z)) = H'_{11}(z) \cdot H'_{12}(z) - H'_{21}(z) \cdot H'_{22}(z)$$

$$= H_{11}(\frac{\xi}{\gamma}) \cdot H_{12}(\frac{\xi}{\gamma}) - H_{21}(\frac{\xi}{\gamma}) \cdot H_{22}(\frac{\xi}{\gamma}) = \det(H(\frac{\xi}{\gamma}))$$

(3.47)
3.4. Practical problems and solutions

From this, it is clear that applying the spectral smoothing on the individual transfers of the matrix results in a spectral smoothing of the determinant similar to that in the one-channel case. Windowing in the time domain (with $w[k] = \gamma^k$ $k = 0..L$) is equal to a convolution in the frequency domain with lowpass filter, and thus small peaks in frequency response $H(z)$ are widened applying SST, hence the name.

A similar effect can be obtained before LP analysis, by windowing the autocorrelation function.

3.4.4. Regularization of the rotator optimization

For signals with no cross-correlation at lag 0 and equal power, no optimal angle is defined, every angle goes. In practice that means that for slightly different input signals, a totally different angle can emerge. Fortunately, ill-conditioned problems can be detected through the modulation depth; the modulation depth is a measure readily available when determining the optimal angle (see section 3.3).

There are two angles to which biasing is possible. First, from a coding point-of-view, it is most efficient to apply the same angle in the current as in the previous frame. This gives a smoother series of optimal angles over time, and can be exploited through coding the differences between succeeding angles instead of directly coding the angle. However, for quality reasons an angle of $\pm \frac{\pi}{4}$ is preferred at least for uncorrelated nearly equal power input signals because then degradations due to e.g., quantization will have the same effect on both signals, and is probably less annoying then when one channel is highly degraded.

In conclusion, for low modulation depths, biasing is wanted either towards the previous angle or the $\pm \frac{\pi}{4}$ angle.

Another suggestion could be to determine the optimal angle for a filtered signal (e.g., 4 kHz Low-pass) and applying this on the non-filtered signal. This from the background knowledge of intensity stereo: magnitude and phase are dominant spatial cues at the lower frequencies but not at the higher frequencies.

So far, no definite solution has been found and it is recommended to consider these issue more closely.
4. Quantization of the Stereo LPC coefficients

The proposed system uses forward adaptive linear prediction, therefore the LP coefficients need to be quantized, transmitted and stored. The main goal is of course to use the least amount of bits, and by a method which gives the least distortion to the analysis and synthesis characteristics. Direct quantization of the LP coefficients is not really efficient because of the large dynamic range of the filter coefficients, and it is known from the one-channel system that the spectral characteristics are extremely sensitive to such quantization. Also after such a quantization, the stability of the synthesis filter is not guaranteed. Therefore it is better to quantize the LP coefficients in some other representation.

Wanted is a representation with properties such as: parameters with a bounded range and an easy stability checking condition of synthesis filter. This chapter describes, firstly, some known transmission methods for the one-channel case, and secondly, two possible methods are briefly introduced, which perhaps can be applied for the two-channel case.

4.1. Transmission methods for one-channel Linear Prediction

Some of the well-known representations for quantizing one-channel LP coefficients are:

- Reflection Coefficients (RC) and the arcsine representation;
- Log Area Ratio (LAR's);
- Line Spectrum Frequency (LSF's).

**Reflection coefficients**

One way of representing the LP coefficients is by reflection coefficients (RC). This denomination is a result of the interpretation as physical parameters of an acoustical tube. In speech coding, where LP is a much applied tool, an acoustical tube is used to model the vocal tract. The model consists of a series of cylindrical tube sections with different diameters (see Fig 4.1), and can be described by reflection coefficients.

Reflection coefficients have the property that the associated filter is unconditionally stable if all the coefficients are of magnitude less than one. We note that the reflection coefficients are directly available, as intermediate variables, when the Levinson-Durbin algorithm (see appendix B) is used for solving the LP equations. In speech coding, acoustical tubes are used for modelling the vocal tract, and in the digital analog (see Fig. 4.2), the successive sections are specified by the reflection coefficients $K_i$. 
4.1. Transmission methods for one-channel Linear Prediction

Confidential

Figure 4.1.: Acoustical tube model.

Figure 4.2.: Digital analog of a tube section (a), Lattice section (b).

The complete synthesis filter can be implemented as a lattice filter using the sections from Fig. 4.2. This is shown in Fig. 4.3.

The reflection coefficients $K_i$ can be quantized directly. However, using a uniform quantization grid in the range $[-1,1]$ for each coefficient is not optimal in the sense that the sensitivity of the spectrum of the synthesis filter is much larger when the magnitude of $K_i$ is close to one than when it is nearly zero. Therefore, the arcsine function is often employed as a mapping on these reflection coefficients in order to equalize this. It is defined as

$$\Theta_i = \sin^{-1} K_i$$

and, consequently, the value of each $\Theta_i$ is restricted to $[-\pi/2, +\pi/2]$. The mapping is shown in Fig. 4.4. On the new scale, a uniform quantization can then be applied.

Log Area Ratio

Next to the arcsine representation, the Log Area Ratios were proposed to address the same issue: a more uniform sensitivity of the spectral characteristics of the synthesis filter to changes
Chapter 4. Quantization of the Stereo LPC coefficients

Figure 4.3.: IIR Lattice filter implementation for reflection coefficients.

in the parameters. The transformation is defined by

\[ \Gamma_i = \log \frac{1 + K_i}{1 - K_i} \]  

(4.2)

and is depicted in Fig. 4.4.

Figure 4.4.: mapping of reflection coefficients

It is possible to quantize with Scalar or vector quantization. Scalar quantizers quantize each LPC parameter independently, while vector quantizers use a set of LPC parameters. Since the latter is more flexible, it allows more control over the spectral distortion.

**Line spectral frequencies**

Another representation are the Line Spectral Frequencies (LSFs) also known as Line Spectral Pairs (LSPs). In that case, two polynomials \( P \) and \( Q \) are constructed from \( A(z) \) by

\[
P(z) = A(z) + z^{-(N+1)} A(z^{-1}),
\]

(4.3)

\[
Q(z) = A(z) - z^{-(N+1)} A(z^{-1}).
\]

(4.4)
4.2. Quantization of two-channel prediction parameters

If \( A(z) \) is minimum phase polynomial, then all the roots of \( P(z) \) and \( Q(z) \) lie on the unit circle. Furthermore, the roots of \( P(z) \) and \( Q(z) \) are interspersed. Thus, the LSFs are ordered and bounded. There is also a physical interpretation possible of these two polynomials [29]: \( P(z) \) corresponds to the vocal tract with the glottis closed and \( Q(z) \) with the glottis open.

4.2. Quantization of two-channel prediction parameters

For two-channel LP systems, the quantization of the filter coefficients is an unknown territory. The one-channel approaches described before do not apply immediately: they have to be adapted or extended in order to be applicable to the two-channel case. We describe two preliminary ideas on the quantization. We do not consider direct quantization of the matrix polynomial coefficients because, from the experience in the one-channel case, we assume that is about the worst you can do.

4.2.1. Reflection Matrices

Since the two-channel case is in so-far identical to the one-channel case that a (block-)Levinson algorithm applies, and since in the one-channel case both the arcsine and the LAR representation are coupled to the reflection coefficients, it seems worthwhile to consider the reflection matrices as arising in the block-Levinson algorithm as a possible source for quantization strategies.

A first problem that appears here is that, in contrast to the one-channel case, two sets of reflection matrices appear in a block-Levinson algorithm: one for the forward and one for the backward recursion. Having either one of them is insufficient to start the recursion. Therefore, transmission of the reflection matrices does not seem to be cost effective. Fortunately, both forward and backward reflection matrices can be derived from the so-called normalized reflection matrices. To do this, data concerning one additional matrix has to be transmitted. This data is then iteratively updated in order to be able to construct both reflection matrices from the normalized one.

A possible way for transmitting a (normalized) reflection matrix could be the following. As any matrix, it can be decomposed by either a singular value decomposition or an eigenvalue decomposition. In both cases, we obtain 2 rotation angles (coupled to the unitary matrices or the eigenvectors) and the singular values or eigenvalues. It is believed that transmitting these singular values or eigenvalues in the form of and arcsine or LAR presentation could be a possible way. Of course, it would also be possible to code both the eigenvalues in some (arithmetic/geometric) mean and difference value. Furthermore, the angles coupled to the vectors can be quantized as well, with an accuracy possibly dependent on the strength of its associated singular or eigenvalue.

4.2.2. Coupling to OCS parameters

In view of the experience and knowledge obtained with quantizing the Interaural Time Difference (ITD), Interaural Level Difference (ILD) and coherence values in OCS (subsection 2.2.6), it would be advantageous to have a coupling between these parameters and the parameters of a two-channel LP system. Although the description of the LP system in terms of an all-pass line
Confidential

Chapter 4. Quantization of the Stereo LPC coefficients

(3.13) suggest that that is not impossible since the all-pass is a warping device and, consequently, DFT-transforms act like a uniform sampling on a perceptual frequency (Bark) scale (just like OCS), the relation between the OCS parameters and the two-channel LP parameters is certainly not straightforward. To start with, the LP system uses 4 parameters per section, where one could say that OCS works with 3 parameters per section. The extra fourth parameter in the LP system has probably to do with the overall whitening of both channels, an effect not taken into account in OCS.

So far, no research has been carried out on this specific topic. This is recommended because presumably this does not only lead to a method of quantization, but can also deepen our insights in both systems.
4.2. Quantization of two-channel prediction parameters Confidential
5. Signal quantization experiments

Next to the parameters (LP coefficients and rotation angle), other necessary information which needs to be transmitted or stored for reconstruction, are the rotated residual signals (main and side signal). If the proposed system is to compete with the state-of-the-art audio coders, it is also necessary to quantize these signals as efficiently as possible.

The bits spent on the parameters (or side information) are just assumed to be a small fraction of the total bit rate, because these parameters are transmitted only once per frame. The overall bit rate is therefore mainly determined by the costs of transmitting the residual signals. Therefore, several experiments were conducted to establish which parts or components of these main and side signals are essential in order to maintain a good quality of the reconstructed signal, and which parts can be compromised. These experiments (sections 5.2-5.7) were done with a collection of divers excerpts (see appendix A.3), representing different musical and speech(-like) signals.

Before discussing these experiments, we will first show the functioning of the system through an example (section 5.1).

Note that the experiments do not aim at removing psycho-acoustical irrelevancies, the purpose of the conducted experiments is examining how removing parts of the main and side signal affects the output. For removing the irrelevancies more insight into the proposed system and the inclusion of an appropriate psycho-acoustical model is needed.

5.1. Signals in the system

By means of plotting the different intermediate signals, we demonstrate the operation of the proposed analysis system. This is done for a part of an excerpt and with a sufficiently high prediction filter order, but similar results are achieved for other excerpts. Furthermore, we consider the estimated synthesis system.

Analysis system

The input signal, used for this example, is spread over multiple frames. It consists of both tonal and noise-like parts (Fig. 5.1(a)), and has a non-flat spectrum (Fig. 5.1(b)). Moreover, some similarities between both channels can be detected in the time domain.

The residual signals ($e_1$ and $e_2$), from which the predicted components have been removed, are shown in Fig. 5.2. The tonal parts of the input are almost completely removed (Fig. 5.2(a)), because of the high predictability of these parts. Furthermore, a clear drop in amplitude can be seen (cf. scales of y-axis in Fig. 5.1(a) and 5.2(a)), not only for the tonal parts but also for the
more noise-like parts. This results in a drop of total energy and, in addition, energy is evenly
distributed resulting in a nearly flat spectrum (Fig. 5.2(b)), because of redundancy removal.

Finally, possible correlation between sample pairs is removed through the rotator. This
correlation at lag zero can still exists, because the LP only makes an estimates according to the
history of both input signals, but not from the current samples. Therefore, after rotation we
see a lower amplitude for the side signal, and a slight increase for the main signal (Fig. 5.3(a)).
Furthermore, the spectrally flatter character is still maintained after rotation (Fig. 5.3(b)).

![Figure 5.1.: Input signal in: Time domain (a), Frequency domain (b).](image)

![Figure 5.2.: Residual signals in: Time domain (a), Frequency domain (b).](image)

Now, for the same example, an additional explanation is given according to Lissajous plots,
in these plots the current sample of one channel is plotted against the current sample of the
other channel for different instances in time. The Lissajous plot of the input (Fig. 5.4(a)) shows
a strong dependence between both channels (here the mono component). Next, the residual
signal, which results after LP, has a much lower variance than the input (Fig. 5.4(b)), but
the dependence between the sample pairs of the channels still exists, with almost the same
orientation as the input. Therefore, the rotator is used for removing this last correlation. The
Confidential

Chapter 5. Signal quantization experiments

Figure 5.3.: Main and side signals in: Time domain (a), Frequency domain (b).

rotation of the $e_1$-$e_2$-plane can be clearly seen in Fig. 5.4(c). For this excerpt it results in a clear main and side signal, which are completely uncorrelated.

Figure 5.4.: Lissajous plot of: input (a), residual (b), rotated residual (c).

**Synthesis filter**

The transfer of the 2-channel synthesis filter, given by eq. (3.7), can be characterized through an eigenvalue decomposition of the sampled transfer matrix $H(e^{-j\theta})$

$$H(e^{-j\theta}) = \begin{bmatrix}
1 - P_{11}(e^{-j\theta}) & -P_{12}(e^{-j\theta}) \\
-P_{21}(e^{-j\theta}) & 1 - P_{22}(e^{-j\theta})
\end{bmatrix} = \mathbf{V}(\theta)\mathbf{E}(\theta)\mathbf{V}(\theta)^{-1}
$$

(5.1)

with $\theta$ uniformly sampled on the unit circle from 0 to $\pi$ (equivalent to $[0, \frac{\pi}{2}, F_s]$). An eigenvalue decomposition is performed per frequency, which results in two (possibly) complex eigenvalues. For characterizing the synthesis filter the inverse magnitude of these eigenvalues are needed. Therefore, in the following figures the input spectrum of both channels, and the inverse magnitudes of the two eigenvalues are plotted on a dB-scale, for a system using a TDL and spectral
5.2. Uniform quantization

This and the next sections concern experiments with quantization of the main and side signals, to determine the relevant components of these signals. Because of the lower variance of the main and side signal it is possible to quantize them more efficiently, and therefore less bits are needed for these signals.
The first section describes an experiment conducted with uniform quantization, to determine the number of quantization levels needed for transparent coding. To do so, the signal is chopped into equal length segments. Next the segment is normalized between \([+1, -1]\) and finally, this normalized segment is quantized with a certain number of levels. The quantized signal is then transmitted to the decoder, with an additional normalization factor per segment. For audio coding it is recommended to always include the zero level, in order to preserve silences, and to prevent alternating quantization of background noise. Therefore an odd number of levels is used.

The first experiment is intended for determining the number of quantization levels needed for the side signal, which gives a transparent output, while maintaining the original main signal. The number of levels ranged from 3 till 11. The transparency was tested through an informal listing test with headphones.

First the side signal was quantized with only 3 levels (with levels \([-i; 0; +i]\) for a normalized segment). This resulted in a fair reconstruction for only few mono-like excerpts. But for most of the excerpts, with less correlation between the channels, it resulted in a rough/coarse signal. Also input signals containing very tonal parts sounded after reconstruction as if they were modulated (slightly vibrating tones).

Next, the outcome for a 5-level quantization resulted in a fair reconstruction for most of the excerpts. Only one excerpt containing a cello still sounded modulated.

For 7 levels and more all but one excerpt were considered transparent. Only an excerpt, containing applause, still sounded dull, with onsets are not perceived that strong. The strong onsets in this excerpt, which still remain in the side signal after LP, leads to a kind of pre-echo-like artifact after reconstruction. But this probably can be solved with a more efficient segmentation (window length switching).

Concluded from this experiment is that at least 7 quantization levels are needed for achieving transparent coding, while still leaving the main signal intact. Results in at least 3 bits per sample only for the side signal, which is too much for achieving low bit rates.

### 5.3. Nonuniform quantization

Because uniform quantization still needs too many bits another experiment is conducted, to determine the number of levels needed when using nonuniform quantization. Uniform quantization is most efficient for signals with a uniform distribution. But the distribution of the side signal is not uniform (it is, in fact strongly peaked around zero).

A rough approach for the probability function is also implemented. First, the signal is again chopped into segments and normalized (Fig. 5.7(a)), but next for determining the decision and quantization levels the data is sorted (Fig. 5.7(b)). The x-axis is divided into the same number of pieces as the required quantization levels. Next the decision and quantization levels are alternatingly determined by the value of the sorted values at these division points. Figure 5.7(b) shows this division for 5 levels, the dotted lines are the decision levels and the solid straight lines are the quantization levels. This gives a rough estimate of the probability density function. The quantized frame is shown in Fig. 5.7(c), this differs form uniform quantization where quantization levels \([-\frac{4}{5}, -\frac{2}{5}, 0, \frac{2}{5}, \frac{4}{5}]\) are used. More side information is needed when using nonuniform
5.4. Side signal removal

Besides the normalization factor also the actual levels have to be transmitted (half of the levels because of symmetry around the zero level).

![Normalized frame](a), sorted absolute values (b), quantized frame (c).

This experiment resulted in almost the same outcome as the previous, with some improvements for a few excerpts, but with degradation in others (onsets for some excerpts are not as strong). Non-uniform quantization clearly improved the quality of tonal excerpts, but the modulation is not completely removed. Overall still 7 levels are needed for the side signal. Using a more efficient segmentation probably gives a slight increase in quality. But these improvements have to be balanced against a more complex quantization method with extra side information.

**5.4. Side signal removal**

For getting a better insight into the perceptual relevance of the side signal, an experiment is conducted where the side signal is completely discarded. Applying rotation on a signal with a strong dependence between the sample pairs, results in a side signal with a very low energy (and thus little information) compared to the main signal. Theoretically, discarding this side signal should not result in a big loss of quality, however through this experiment also the result for more uncorrelated signals is examined.

The difference in energy between the main and the side signal is, for mono files about 40 dB. Discarding the side signal still results in a good reconstruction. Although for one monolike excerpt, Suzanne Vega, this still resulted in a reconstruction with a low frequency artifact (noise). The difference between main and side energy is about 20 dB, but taking a closer look at the spectrum of the main and side signal it was discovered that not all frequencies had such huge difference. For the very low frequencies (0-10 Hz) the energies are equal. The reason for this is unclear.

Furthermore, for almost all of the other excerpts the disposed side signal resulted in an unstable stereo image, where the source positions swapped between the channels, and it sounded a bit coarse.

The causes for these artifacts are not completely understood. However it is clear that we are not allowed to do rather arbitrary operations on the side signal. There are still meaningful components in the side signal which have to be maintained.
5.5. Synthetic side signal

As described in the outcome of the previous experiment, the side signal still contains information on the position of the source, and this can not be discarded. Therefore, another experiment is conducted where a synthetic side signal is generated at the decoder side, to compensate for the loss in energy resulting after discarding the side signal.

Generating the synthetic side signal is done similar to OCS and VECMA. It is made by delaying the main signal (about 500 samples), to decorrelate it with the main signal (originally main and side signals are orthogonal) (Fig. 5.8). Next, the total energy of the frame is corrected, according to a scale factor determined at the encoder, for which one extra parameter is needed.

Again an informal listening test was used for determining the quality of this method. The synthetic side signal resolved the unstable stereo image issue. The source positions are now maintained, but it sounded very noisy, i.e. low frequency noise (below 250 Hz), for all the non-mono excerpts.

Figure 5.8.: Create Side out of delayed main signal.

5.6. Partial side signal transmission

Considering the previous two experiments, which both have low frequency artifacts, a follow up experiment is tried. Now only the low frequency (3-4 kHz) part of the side signal (LF-side) is transmitted, omitting everything else. This resulted in a stable stereo image but still a somewhat noisy reconstruction. But the main problem, i.e. an unstable stereo image, which occurred when removing the side signal completely, were absent. Furthermore, the reconstruction is less noisy than with a synthetic side signal. This implies that most of the meaningful information of the side signal is in the low frequency part.

Next, this low frequency side signal is combined with the 'missing' high frequency part of the synthetic side signal. This resulted in an overall improvement, and led to fair results where most of the noise was removed.

The conclusion from these experiments is that a reduction to one channel is not directly possible, but fair result can been achieved with the original LF-side signal combined with the synthetic signal, which is an equivalent to a 1.1-channel system (in bandwidth) instead of a 2-channel.
5.7. Spectral bandwidth replication

As done in speech coding, and more recently also applied in audio coding [11], we applied Spectral Bandwidth Replication (SBR) for reducing the total bandwidth.

With SBR (used in HE-AAC and MP3Pro) only the lower frequency part, i.e. half of the bandwidth, is transmitted, because of the lower psycho-acoustical importance of the temporal signal structure at high frequencies. In the decoder the reconstruction of the high frequency part is done by mapping the low frequency part to the missing high frequencies. Then these are shaped through some scaling data, in order to match the original data and finally some tonal and noise-like parts can be added. This bandwidth reduction makes it a very useful tool for applications which require low bit rate audio coding.

A sufficiently high prediction order used in the proposed scheme, results in a residual signal with a spectrally flat envelope, and which is still maintained after rotation (Fig. 5.9(a)). Because of this signal property it is much easier to apply SBR, making it less complex. The main and/or side signal is downsampled with a decimation factor of 2 and 3th order Chebyshev filter, saving half the bandwidth. At the decoder the signals are upsampled again, resulting in a mirrored version of the lower part at the higher frequencies (Fig. 5.9(b)). No additional side information is needed because of the spectrally flat nature of the main and side signal. There is only a small aliasing difference at the highest frequencies, but this will be shaped through the synthesis filter.

Two experiments were conducted, first applying SBR on the side signal only, and next applying SBR on both side and main signal, generated using prediction order 20.

Only SBR on the side signal results in almost transparent reconstructions for most of the excerpts. Only a small difference was perceived for the castanets excerpt, in which some high frequency noise can be heard. Applying SBR on both main and side resulted a slight degradation for some of the excerpts (compared to the previous experiment), e.g. in the castanets more high frequency noise is introduced.

Main conclusion is that the quality is good, although not completely transparent for all excerpts.
5.8. Conclusion from the experiments

From the experiments we conclude that the side signal can not be discarded completely. There are still meaningful components in the side signal which have to be maintained, but other experiments indicate that most important components are within the lower frequency region.

The most promising technique for attaining a low bit rate is bandwidth reduction through SBR, which can be easily used on the spectrally flat signals. Furthermore, it can be applied on the side signal as well as on the main signal, resulting in the equivalent bandwidth of one mono channel. Overall, we can probably do slightly better than OCS/BCC, because the side signal can be reduced even further, and this suggests that low bit rate coders based on the proposed system are feasible.
5.8. Conclusion from the experiments
6. Conclusions and Recommendations

6.1. Conclusions

The aim of this thesis was to explore a stereo coding technique capable of providing similar or better performance, in quality/bit rate balance, than existing coders, and for both the high and the low bit rate levels. It was argued that known stereo coding techniques do not simultaneously deliver the best performance at both bit rate levels.

Considered was a 2-channel Linear Prediction (LP) plus rotator, since this allows perfect reconstruction (which is in line with the best quality at high bit rate end) and allows the construction of a main and side signal (similar to OCS; the best low bit rate stereo coding tool).

For the proposed system it was shown that the optimization dissolves in two separate stages: optimizing the stereo LP and subsequently optimizing the rotator with the residual from the Stereo LP. For the prediction filters Laguerre filters are chosen, because of their close resemblance with the psycho-acoustical Bark-scale, which is advantageous for lossy audio coding. For equal cross and auto prediction orders the block Levinson algorithm applies, which is an efficient way of optimizing the Stereo LP. The block Levinson algorithm implies that the synthesis system is stable, for equal orders, and that the required order can be adaptively determined.

This system is implemented and tested. The tests show that the implemented system operates as expected. Special care was needed for some pathological input signals.

A number of informal listening experiments were performed to establish which components of the residual signal have to be maintained in a transmission system and which parts can be compromised. Discarding the side signal completely gives problems for most cases, but fair results are already achieved by only transmitting 10% of the side signal (only low frequencies). But in view of the spectral flattening, Spectral Band Replication (SBR) appears to be a method which allows low bit rate coding for the proposed system.
6.2. Recommendations

On basis of the research reported here and in line with the conclusions given above, we have the following list of recommendations:

- Regularization needs to be considered in more detail, this thesis only describes preliminarily solutions.
- Efficient quantization techniques for the parameters have to be developed, two possible directions for this are suggested within this thesis.
- To get a better insight into the required bit rate for the transmission of the residual signal, more experiments are required. This in order to pin down the essential components of the residual signal.
- Finally, the total complexity has to be estimated and compared with the existing coding systems.
A. Lists of abbreviations, notations and used excerpts

A.1. List of abbreviations

BCC : Binaural Cue Coding
CELP : Code Excited Linear Prediction
DFT : Discrete Fourier Transform
FFT : Fast Fourier Transform
FIR : Finite Impulse Response
HE-AAC : High Efficiency Advanced Audio Coding
ICLD : Interaural Channel Level Differences
ICTD : Interaural Channel Time Differences
IID : Interaural Intensity Differences
IIR : Infinite Impulse Response
ILD : Interaural Level Differences
ITD : Interaural Time Differences
LAR : Log Area Ratios
LP : Linear Prediction
LPC : Linear Predictive Coding
LSF : Line Spectral Frequencies
LSP : Line Spectral Pairs
LWRR : Levinson Wiggins Robinson algorithm
OCS : Optimum Coding of Stereo
PCA : Principle Component Analysis
PCM : Pulse Code Modulation
SBR : Spectral Bandwidth Replication
SVD : Singular Value Decomposition
TDL : Tapped Delay Line
VECMA : Very Efficient Coding of Multichannel Audio
A.2. List of notations

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>a, b, c</td>
<td>Lowercase letters are signals in the time domain</td>
</tr>
<tr>
<td>A, B, D</td>
<td>Uppercase letters are signals in the frequency domain</td>
</tr>
<tr>
<td>x[n]</td>
<td>Discrete time index n of signal x</td>
</tr>
<tr>
<td>X(z)</td>
<td>Frequency index z of signal X</td>
</tr>
<tr>
<td>ť</td>
<td>Predicted signal of x</td>
</tr>
<tr>
<td>ź</td>
<td>Reconstructed signal of x</td>
</tr>
<tr>
<td>A, B, C</td>
<td>Bold uppercase letters are matrices</td>
</tr>
<tr>
<td>ã, b, ć</td>
<td>Underlined lowercase letters are vectors</td>
</tr>
<tr>
<td>x₁</td>
<td>Left channel of stereo signal</td>
</tr>
<tr>
<td>x₂</td>
<td>Right channel of stereo signal</td>
</tr>
<tr>
<td>A^{(m)}</td>
<td>Matrix A in a loop at iteration m</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Operator</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>*</td>
<td>convolution</td>
</tr>
<tr>
<td>c*</td>
<td>Complex conjugated of c</td>
</tr>
<tr>
<td>det(A)</td>
<td>Determinant of matrix A</td>
</tr>
<tr>
<td>A⁻¹</td>
<td>Inverse of matrix A</td>
</tr>
<tr>
<td>Aᵀ</td>
<td>Transpose of matrix A</td>
</tr>
<tr>
<td>A*</td>
<td>Complex conjugate of matrix A</td>
</tr>
<tr>
<td>Aᴴ</td>
<td>Conjugate transpose of matrix A</td>
</tr>
</tbody>
</table>

A.3. List of used excerpts

The following excerpts are used, all wav-files with a sample frequency of 44.1 kHz.

<table>
<thead>
<tr>
<th>Title</th>
<th>Content</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suzanne Vega - Tom’s Diner</td>
<td>Voice (mono-like)</td>
<td>0:10</td>
</tr>
<tr>
<td>German male</td>
<td>Voice (mono-like)</td>
<td>0:08</td>
</tr>
<tr>
<td>Speech English Female</td>
<td>Voice (mono-like)</td>
<td>0:07</td>
</tr>
<tr>
<td>Trumpet</td>
<td>Single instrument</td>
<td>0:10</td>
</tr>
<tr>
<td>Orchestra</td>
<td>Multiple instruments</td>
<td>0:12</td>
</tr>
<tr>
<td>Quiz Tune</td>
<td>Multiple instruments</td>
<td>0:11</td>
</tr>
<tr>
<td>Harpsichords</td>
<td>Single instrument</td>
<td>0:07</td>
</tr>
<tr>
<td>Castanets</td>
<td>Single instrument</td>
<td>0:07</td>
</tr>
<tr>
<td>Harmonica</td>
<td>Single instrument</td>
<td>0:27</td>
</tr>
<tr>
<td>Bag pipes</td>
<td>Single instrument</td>
<td>0:11</td>
</tr>
<tr>
<td>Xylophone</td>
<td>Single instrument</td>
<td>0:10</td>
</tr>
<tr>
<td>Piano</td>
<td>Single instrument</td>
<td>0:13</td>
</tr>
<tr>
<td>Applause</td>
<td>Applause</td>
<td>0:10</td>
</tr>
<tr>
<td>Pop</td>
<td>Multiple instruments</td>
<td>0:10</td>
</tr>
<tr>
<td>Byrds - Eight Miles High</td>
<td>Multiple instruments</td>
<td>0:33</td>
</tr>
<tr>
<td>Carmen - L’amour est un oiseau rebelle</td>
<td>Opera</td>
<td>0:10</td>
</tr>
</tbody>
</table>
B. One-channel linear prediction

For removing redundancies within an (audio) signal, Linear Prediction (LP) [35, 28] can be used. This because successive samples are often highly correlated. So it should be possible to get a good estimate of the current sample from its preceding ones and thereby to decorrelate the samples. The predicted signal is modelled as a linear combination of its past values by

\[ \hat{x}[n] = \sum_{k=1}^{N} \alpha_k x[n-k] \]  

with the prediction signal \( \hat{x} \), the weights \( \alpha_k \) as linear predictor coefficients and \( N \) is the number of preceding samples being used in the prediction (prediction order). The error which remains after subtracting the predicted signal \( \hat{x}[n] \) from the original \( x[n] \) is called the prediction residual or prediction error \( e[n] \) and is given by

\[ e[n] = x[n] - \hat{x}[n] = x[n] - \sum_{k=1}^{N} \alpha_k x[n-k]. \]  

Such a system is called an one-step ahead predictor because it tries to predict the next sample from its previous ones. This results in an all-zero prediction filter, of order \( N \) (see Fig. B.1), and with an impulse response \( a[k] = -\alpha_k \) (\( k = 1..N \) with \( a[0] = 1 \)). The corresponding transfer function of the analysis filter is given by

\[ A(z) = \sum_{k=0}^{N} \alpha_k z^{-k} \]  

With the correct predictor coefficients (the computation of these optimal parameters will be explained later on), the short-term correlations between samples of the residual have been minimized, and this leads to an approximately flat spectral envelope for the residual, that is if a sufficiently high order is used. Therefore, the analysis filter is also known as a whitening filter.

The variance of the residual is lower (or equal for a signal with no correlation) than the original, therefore it can be quantized more efficiently. However, this is signal dependent, because some signals are more predictable than others. This ratio in variances of the residual and the input signal is called the prediction gain, and is defined as

\[ G = \frac{\sigma_x^2}{\sigma_e^2} \quad \text{or} \quad G_{\text{dB}} = 10 \cdot \log_{10} \left( \frac{\sigma_x^2}{\sigma_e^2} \right) [\text{dB}]. \]
For reconstructing the original, a synthesis filter (Fig. B.2) is used. It consists of the same predictor, but now in a feed-back loop. The synthesis system is an all-pole filter with transfer

$$H(z) = \frac{1}{A(z)} = \frac{1}{\sum_{k=0}^{N} \alpha_k z^{-k}}. \tag{B.5}$$

For guaranteeing a stable synthesis filter, $H(z)$ must have all its poles with the unit circle and, because the synthesis filter is the inverse system of the analysis filter, the same is demanded of the zeros $A(z)$.

In the absence of quantization this system is capable of perfect reconstruction (Fig. B.3), i.e. the reconstructing is a bit true copy of the input. Therefore, LP can also be a useful tool for lossless coding.
Calculation of the optimal parameters

The optimal predictor coefficients are determined by minimizing the total squared error

\[
E = \text{min}\left\{ \sum_{n=1}^{N} e^2[n] \right\} = \text{min}\left\{ \sum_{n=1}^{N} (x[n] - \hat{x}[n])^2 \right\} = \sigma_e^2
\]  

(B.6)

For the minimum of \( E \) holds that the derivatives are equal to 0:

\[
\frac{\partial \sigma_e^2}{\partial \alpha_k} = 0, \quad 1 \leq k \leq N.
\]  

(B.7)

This yields to following set of equations (Yule-Walker equations)

\[
\sum_{k=1}^{N} \alpha_k R_{k-i} = R_i, \quad 1 \leq i \leq N.
\]  

(B.8)

where \( R_k \) is the \( k \)th autocorrelation coefficient of the input signal and is given by

\[
R_k = \sum_{i=k}^{N} x[i]x[i-k].
\]  

(B.9)

Note that \( R_k \) is an even function of \( k \) (for real-valued input), i.e.

\[
R_k = R_{-k}.
\]  

(B.10)
Equation B.8 can be rewritten in a matrix notation

\[ R_{\alpha} = r \]  

with the \((N \times N)\) autocorrelation matrix \( R \) given by

\[
R = \begin{pmatrix}
    R_0 & R_{-1} & R_{-2} & \cdots & R_{-N+1} \\
    R_1 & R_0 & R_{-1} & \cdots & R_{-N+2} \\
    R_2 & R_1 & R_0 & \cdots & R_{-N+3} \\
    \vdots & \vdots & \vdots & \ddots & \vdots \\
    R_{N-1} & R_{N-2} & R_{N-3} & \cdots & R_0
\end{pmatrix}
\]  

(B.12)

which has a Hermitian Toeplitz structure and in the case of real-valued input signals a symmetric Toeplitz structure. The optimal prediction coefficients are stacked into a vector \( \alpha = [\alpha_1, \alpha_2, \cdots, \alpha_N]^t \), and \( r \) is also a stacked vector \( \alpha = [R_1, R_2, \cdots, R_N]^t \). The optimal prediction coefficients can be computed by solving B.11, this results in

\[ \alpha = R^{-1} r. \]  

(B.13)

Using the autocorrelation method this requires \( \frac{N^2}{2} + O(N^2) \) operations (multiplications and divisions)[35].

However, by making use of the structure of the matrix \( R \) (symmetric Toeplitz) it is possible to solve B.13 in a more efficient manner. Like in the recursive Levinson-Durbin algorithm which also makes use of the fact that the vector \( r \) comprises the same elements as in the autocorrelation matrix \( R \). The algorithm can be specified as follows

\[
E_0 = R_0
\]

(B.14a)

\[
k_i = \frac{R_i + \sum_{k=1}^{N} \alpha_j^{i-1} R_{i-j}}{E_{i-1}}
\]

(B.14b)

\[
\alpha_j^{(i)} = k_i
\]

(B.14c)

\[
\alpha_j^{(i)} = \alpha_j^{(i-1)} + k_i \alpha_{i-j}^{(i-1)} \quad 1 \leq j \leq i - 1
\]

(B.14d)

\[
E_i = (1 - k_i^2) E_{i-1}
\]

(B.14e)

Equations B.14 b-e are solved recursively for \( i = 1, 2, \cdots, N \), where \( ^{(i)} \) denotes the iteration step, and the intermediate \( k_i \) are also known as reflection coefficients. This method requires only \( N^2 + O((N)) \) operations and, because also all solutions are computed for all orders less than \( N \), it also possible to adaptively determine the order, by incorporating other stop conditions, e.g. dependent on the energy \( E_i \).
IIR based linear prediction

Linear predictive coding (LPC) has been widely used in speech signal processing since its introduction in the late 1960s. This because the transfer function of a lossless tube, used to model the vocal tract, can be described by an all-pole model. Because of the success in speech coding, it is worthwhile to consider if it is applicable to wide band audio coding [38, 32], but here for some adaptations are needed.

For audio coding purposes, it is advantageous to use filters with a behavior that in someway resemble the human auditory filters. This can be done by replacing the tapped delay-line of the normal LPC scheme by IIR filters, e.g. Kautz filters [9], Laguerre filters [8], or all-pass sections like in Warped-LP [22]. Laguerre and WLP effectively results in a frequency dependent delay-line, with which more is emphasized on the lower frequencies of the audio signal. Therefore, a "linear" estimate can be made with a longer history for the lower frequencies, which are psycho-acoustically more important, this of course at the expense of the higher frequencies. For instance, the Pure Linear Prediction (PLP) scheme uses Laguerre filters, which are used to provide a whitening on a Bark-like scale [39]. This effect is demonstrated in Fig. B.4, where the spectrum of the input signal is plotted (input data) together with the synthesis filter transfer for LP and WLP. We observe that WLP has more details in the low-frequency side of the spectrum, while LP provides more (psycho-acoustically unnecessary) details at the high-frequency side.

Figure B.4.: Differences between warped and normal LP.
C. Rotation

A basic building block in the proposed system is the rotator. Rotation, main axis transform, applying Singular Value Decomposition (SVD) or Principle Component Analysis (PCA) are all identical transformations, all with the same intention namely to minimize the covariance between sets of data.

Rotating over an angle $\varphi$ (counter clockwise around the origin) in a two-dimensional plane is defined by multiplying the data sets with the transformation matrix $R$

$$R(\varphi) = \begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix}. \quad (C.1)$$

Note that the inverse rotation is equal to the rotation with the negative angle

$$R(\varphi)^{-1} = R(-\varphi). \quad (C.2)$$

For audio coding, rotation is used for removing correlations between the current samples of the two channels (Fig. C.1), which results in a main and a side signal.

For the mono-like input signal, shown in figure C.2(a), the strong dependence between channels can be easily detected in a Lissajous plot (in here current values of $x_1$ are plotted against current values of $x_2$ for different instances in time) (Fig. C.3(a)). After rotation it results in a main signal ($m[n]$) with the mono component, and a low energy side signal with the "differences" of both channels (C.2(b)). The second Lissajous plot (Fig. C.3(b)) shows clearly the rotation of the $x_1 - x_2$-plane, and also the orthogonality (uncorrelated) of the main and side signal.
Determining the optimal rotation angle

Determining the optimal rotation angle $\varphi$, can be defined as an optimization problem which produces a maximum of a squared sum of the main signal $\max \{ \Sigma_n |e_1[n]|^2 \}$, which automatically induces a minimum for the squared sum of the side signal $\min \{ \Sigma_n |e_2[n]|^2 \}$.

First we define the input vectors by

$$\begin{align*}
\varepsilon_1 &= [ e_1[n], e_1[n + 1], e_1[n + 2], ..., e_1[n + L - 1] ]^T \\
\varepsilon_2 &= [ e_2[n], e_2[n + 1], e_2[n + 2], ..., e_2[n + L - 1] ]^T
\end{align*}$$

and the energies and cross-energy of both both channels are defined by

$$\begin{align*}
R_{11} &= \varepsilon_1^* \cdot \varepsilon_1 \\
R_{22} &= \varepsilon_2^* \cdot \varepsilon_2 \\
R_{12} &= \varepsilon_1^* \cdot \varepsilon_2 = \varepsilon_2^* \cdot \varepsilon_1
\end{align*}$$
From the definition
\[ J = \sum_n (e[n])^2 \] (C.6)
with (C.4) and (C.5), \( J \) can be rewritten to:
\[ J = \frac{R_{11} + R_{22}}{2} + \frac{R_{11} - R_{22}}{2} \cos(2\varphi) + R_{12} \sin(2\varphi). \] (C.7)

Since \( J \) is a periodic function with periodicity \( \pi \), we infer the following:
1. the maximum of \( J \) is attained for values \( \hat{\varphi} + k\pi \) \((k \in \mathbb{N})\);
2. the minimum of \( J \) is attained for values \( \hat{\varphi} + k\pi \) \((k \in \mathbb{N})\);
3. the minimums and maximums are \( \pi/2 \) apart: \( \hat{\varphi} = \hat{\varphi} + k\pi/2 \).

In particular, we have
1. \( J_0 = \frac{1}{2}(R_{11} + R_{22}) \) is the mean value over all \( \varphi \);
2. \( J_1 = \frac{1}{2}(R_{11} + R_{22}) + \sqrt{(R_{11} - R_{22})^2/4 + R_{12}^2} \) the maximum of \( J \);  
3. \( J_2 = \frac{1}{2}(R_{11} + R_{22}) - \sqrt{(R_{11} - R_{22})^2/4 + R_{12}^2} \) the minimum of \( J \).

We note that there is always a maximum and minimum except when \( \sqrt{(R_{11} - R_{22})^2/4 + R_{12}^2} = 0 \). We define the modulation depth \( d \) as the ratio of the mean \( J_0 \) and the excursion from the mean:
\[ d = 2 \frac{\sqrt{(R_{11} - R_{22})^2/4 + R_{12}^2}}{R_{11} + R_{22}}, \] (C.8)
and \( 0 < d < 1 \). If \( d \approx 0 \), then action has to be taken to guarantee a well-defined solution.

Define
\[ c = \frac{R_{11} - R_{22}}{2} + jR_{12} \] (C.9)
and
\[ \phi = \angle c \] (C.10)

We then have
\[ J = J_0 + |c| \cos(2\varphi - \phi) \] (C.11)
and thus
\[ \hat{\varphi} = \phi/2 \ \ (\pm k\pi), \] (C.12a)
\[ \hat{\varphi} = \phi + \pi/2 \ \ (\pm k\pi). \] (C.12b)

To remove ambiguity, we take \( \hat{\varphi} = \hat{\varphi} + \pi/2 \).
D. Relation between Laguerre and tapped-delay-line LP systems

In this appendix, it is shown that a two-channel LP Laguerre system is theoretically equivalent to a LP TDL system with prefiltering. The consequence of this equivalence is the following. Assume that for all input signals and under certain optimization conditions, the LP TDL system yields stable synthesis filters. Then, under the same optimization conditions, the optimized LP Laguerre system will do so as well.

The equivalence can be shown as follows. Consider the predictor transfers

\[ P_{ij}(z) = z^{-1} \sum_{k=1}^{N} \alpha_{ij,k} \frac{\sqrt{1-\lambda^2}}{1-\lambda z^{-1}} \left( \frac{-\lambda + z^{-1}}{1-\lambda z^{-1}} \right)^{k-1}. \]  

We can rewrite this to

\[ P_{ij}(z) = \frac{z^{-1} \sum_{k=1}^{N} \alpha_{ij,k} \sqrt{1-\lambda^2} (-\lambda + z^{-1})^{k-1} (1-\lambda z^{-1})^{N-k}}{(1-\lambda z^{-1})^N} = \frac{z^{-1}Q_{ij}(z)}{(1-\lambda z^{-1})^N} \]  

where \( Q_{ij} \) is a \( N-1 \)st order polynomial. In view of the \( N \) free coefficients \( \alpha_{ij,k} \), the polynomial \( Q_{ij} \) is an arbitrary polynomial of order \( N-1 \).

Consider now the transfers \( H_{ij} \).

- For \( i \neq j \) we have \( H_{ij} = -P_{ij} \) and, therefore, free parameters in the Laguerre cross-predictors is equivalent to free parameters in a TDL cross-predictor when operating on a filtered input signal where the pre-filter is given by the transfer function \( (1-\lambda z^{-1})^{-N} \).

- For \( i = j \) we have \( H_{ii} = 1 - P_{ii} \). Bringing this under a common denominator yields

\[ H_{ii}(z) = \frac{(1-\lambda z^{-1})^N - z^{-1}Q_{ij}(z)}{(1-\lambda z^{-1})^N} = \frac{\tilde{Q}_{ij}(z)}{(1-\lambda z^{-1})^N}. \]  

The polynomial \( \tilde{Q}_{ij}(z) \) is an \( N \)th order polynomial with its constant term equal to one and all other coefficients are free. Also this is consistent with a pre-filtered TDL case.

In conclusion, the optimal Laguerre LP is identical to having a pre-filter running on both channels and performing a TDL-based LP. The latter procedure is however ill-conditioned for typical input signals.

From the previous analysis, the attained increased modelling capability at the low-frequency side of the spectrum for the Laguerre LP system with \( \lambda > 0 \) can be easily understood. The pre-filtering puts more emphasis on the LF part of the spectrum and, therefore, increased coding gain can be achieved when modelling that part accurately.
E. Separability of optimization

Suppose we have two signals \( x_1 \) and \( x_2 \). Furthermore, we have a set of regressor signals \( y_k \), \( k = 1, 2, \cdots, 2K \). Typically these will be filtered versions of \( x_i \), e.g.,

\[
\begin{align*}
y_{2i-1} &= h_i * x_1, \\
y_{2i} &= h_i * x_2,
\end{align*}
\]

where \( x_1 \) and \( x_2 \) are both filtered with the same set of filters \( h_i, i = 1, 2, \cdots, K \). We construct the signal \( \mu \) according to

\[
\mu = \cos(\phi)x_1 + \sin(\phi)x_2 - \sum_{k=1}^{2K} \gamma_k y_k.
\]

Consider the criterion \( J \) with

\[
J = \sum_n |\mu(n)|^2.
\]

We will see later on that there is typically a single maximum and a single minimum of \( J \) as a function of the free parameters \( \phi \) and \( \gamma \) such that when minimising/maximising \( J \) we have in fact constructed a two-port \((x_1, x_2) \rightarrow (m, s)\) with \( m \) and \( s \) the main and side signal.

Consider the maximum and minimum of \( J \). These are called \( J_1 \) and \( J_2 \) and the associated signals are \( \mu_1 \) and \( \mu_2 \):

\[
\begin{align*}
J_1 &= \max_{\phi, \gamma} J, \\
J_2 &= \max_{\phi, \gamma} J,
\end{align*}
\]

and

\[
\begin{align*}
\{\phi_1, \gamma_1\} &= \arg\max_{\phi, \gamma} J, \\
\{\phi_2, \gamma_2\} &= \arg\max_{\phi, \gamma} J.
\end{align*}
\]

The coefficients \( \gamma_k \) for which \( J \) attains its minimum or maximum are given by

\[
Q\gamma = P = \cos(\phi)P^{(1)} + \sin(\phi)P^{(2)}
\]

where

\[
\begin{align*}
Q_{k,l} &= \sum_n y_l(n)y_k^*(n), \\
P_k^{(1)} &= \sum_n x_1(n)y_k^*(n), \\
P_k^{(2)} &= \sum_n x_2(n)y_k^*(n).
\end{align*}
\]
Accordingly, we define
\[
    \gamma^{(1)} = Q^{-1}p^{(1)}, \quad (E.10)
\]
\[
    \gamma^{(2)} = Q^{-1}p^{(2)}, \quad (E.11)
\]
\[
    \gamma = \cos(\phi)\gamma^{(1)} + \sin(\phi)\gamma^{(2)}. \quad (E.12)
\]

Thus, given an \(\phi\), the optimal coefficients \(\gamma\) are known. Note that \(\gamma^{(1)}\) and \(\gamma^{(2)}\) can be determined independently from \(\phi\). We further note that \(\gamma^{(1)}\) and \(\gamma^{(2)}\) can be interpreted as the solutions of the following minimization problems:

\[
    T_1 = \sum_n e_1^2(n)
\]

and

\[
    T_2 = \sum_n e_2^2(n),
\]

respectively, with

\[
    e_1 = x_1 - \sum_{k=1}^{2K} \gamma_k(1) y_k, \quad (E.13)
\]

\[
    e_2 = x_2 - \sum_{k=1}^{2K} \gamma_k(2) y_k. \quad (E.14)
\]

Consequently, both solutions can be considered to stem from the minimization of \(T_1 + T_2\). We note that it is also possible to interpret these coefficients as solutions to the minimization of

\[
    T_0 = \left| \sum_n e_1(n) e_2(n) \right|.
\]

This can be easily shown as follows. First note that minimization of \(T_0\) is equal to minimization of \(T_0^2\). Setting the derivative of \(T_0^2\) equal to zero yields

\[
    \frac{\partial T_0^2}{\partial \gamma^{(1)}} = 2e_1^2 e_2(-Y^t)(x_2 - Y\gamma^{(2)}) = 0, \quad (E.15)
\]

\[
    \frac{\partial T_0^2}{\partial \gamma^{(2)}} = 2e_1^2 e_2(-Y^t)(x_1 - Y\gamma^{(1)}) = 0, \quad (E.16)
\]

where \(Y\) is a matrix containing the signals \(y_k\): \(Y = [y_1, y_2, \ldots, y_{2K}]\). Under the assumption that \(e_1^2 e_2 \neq 0\) (this is only attainable in pathological cases), this leads to

\[
    (-Y^t)(x_2 - Y\gamma^{(2)}) = 0, \quad (E.17)
\]

\[
    (-Y^t)(x_1 - Y\gamma^{(1)}) = 0, \quad (E.18)
\]

i.e, equivalent conditions to (E.10) and (E.11).
Consider now the optimal \( \phi \). We assume real-valued signals. We have

\[
J = (\cos(\phi)x_1^t + \sin(\phi)x_2^t + \gamma^tY^t) \cdot (\cos(\phi)x_1 + \sin(\phi)x_2 + Y\gamma).
\]

Eliminating \( \gamma \) using

\[
\gamma = \cos(\phi)Q^{-1}P^{(1)} + \sin(\phi)Q^{-1}P^{(2)}
\]

leads to

\[
J = \cos^2(\phi)R_{11} + \sin^2(\phi)R_{22} + 2\cos(\phi)\sin(\phi)R_{12}
\]

with

\[
R_{11} = x_1^t x_1 - P^{(1)y}Q^{-1}P^{(1)},
\]

\[
R_{22} = x_2^t x_2 - P^{(2)y}Q^{-1}P^{(2)},
\]

\[
R_{12} = x_1^t x_2 - P^{(1)y}Q^{-1}P^{(2)}.\]

From \( R_{11}, R_{22} \) and \( R_{12} \), the optimal \( \phi \) readily follows (for the maximum and the minimum of \( J \); see later).

We can introduce the auxiliary signals

\[
e_1 = x_1 - Y\gamma^{(1)},
\]

\[
e_2 = x_2 - Y\gamma^{(2)},
\]

and have

\[
R_{11} = e_1^t e_1,\]

\[
R_{22} = e_2^t e_2,\]

\[
R_{12} = e_1^t e_2.\]

Thus, the problem of finding the optimal \( \phi \) and \( \gamma_k \) minimising/maximising \( J \) and of constructing \( e \) can be easily done in four steps:

1. Finding the optimal \( \gamma^{(1)} \) and \( \gamma^{(2)} \),
2. Constructing the signals \( e_1 \) and \( e_2 \),
3. Finding the optimal rotation (PCA) for the signals \( e_1 \) and \( e_2 \),
4. Applying this rotator to these signals in order to construct the signals \( \mu_1 \) and \( \mu_2 \) corresponding to a maximum and minimum of \( J \).

In conclusion, the original problem is now split into two separate stages, each involving a separate optimization. In the first two-channel system, we minimize the powers of \( e_1 \) and \( e_2 \) and absolute value of their cross-power. This results in spectrally flat signals \( e_1 \) and \( e_2 \) and a spectrally flat cross-power (under certain conditions for the filters \( h_k \)). In the second step, the rotator, we rotate the signals \( e_1 \) and \( e_1 \) such that the resulting signals are orthogonal.
F. Matlab & C headers

This appendix contains the headers of files of the following Matlab functions used in the implementation of the proposed system. The stereo synthesis filter is implemented in C and compiled to MEX-function, to speed things up.

function A = opt_lpc_stereo(X,N,pole);
function [x, G, H] = block_lev_stereo(Y,N,pole);
function D=stereo_specsmoothing(alpha,gamma,pole,N);
function [Y,P,states] = stereofilter(X,states,order,param,pole);
function [S, HH] = stereo_spectra(apfile,lambda,freq);
function [alpha,y1,y2,depth,a1] = rota_opt(x1,x2);
function [g,s1,s2]=quant_sig_uni(sig,U,Levels);
function out=nonuniformquant(sig,U,levels);
function [y1, y2] = rotate(x1, x2, alpha);

Headers:

function A = opt_lpc_stereo(X,N,pole);
% Description:
% Calculate the optimal prediction coefficients for a
% stereo predictor
%
% Input:
% X: Windowed stereo signal
% N: Orders of the auto and cross predictors [n11 n12 n21 n22]
% pole: Laguerre filter pole (pole=0 -> TDL; pole=0.7564 -> Bark)
%
% Output:
% A: Calculated prediction coefficients [a11 a12 a21 a22]
% matrix is max(N) x 4 (appended with zeros where necessary)
%
% Author: Teun Selten
% Date: 17 September 2003
% Notes: ---

function [x, G, H] = block_lev_stereo(Y,N,pole);
% Description:
% Calculate the optimal prediction coefficients for a
function D = Stereo_specsmoothing(alpha, gamma, pole, N);
% Description:
% Apply spectral smoothing with vector [1 gamma gamma^2 .. gamma^N]
% on all 4 sets of stereo LP coefficients simultaneous and re-normalizes
% to [1 0 0 1; . . . . ; ...]
% Input:
% alpha: laguerre filter coefficients of stereoLPC coder
% gamma: initial weight of spectral smoothing vector
% pole: laguerre pole
% N: order of stereoLPC coder
% Output:
% D: smoothed stereoLPC laguerre coefficients.
% Author: Teun Selten
% Date: 12 January 2004
% Notes: ---

function [Y, P, states] = stereofilter(X, states, order, param, pole);
% Description:
% Filters the data in vector X with the filter described by
% the vectors in param [a11; a21; a22:] to create the filtered
% data Y. The filter is a auto-cross stereo filter.
% Input:
% X: stereo signal

% stereo predictor
% Input:
% Y: Windowed stereo signal
% N: Order of filter N>=1
% pole: Laguerre filter pole (pole=0 -> TDL; pole=0.7564 -> Bark)
% Output:
% x: Calculated prediction coefficients [a11 a12 a21 a22]
% matrix is N x 4
% Author: Teun Selten
% Date: 3 May 2004
% Notes: ---
% states: initial values of the delay line (number of rows depends on max(order)) [state_11; state_12;] ?=1,2 -> state_11 == state_21
% order: gives the orders of the filter [a11 a21 a12 a22]
% param: the coefficients of the filters [a11; a21; a12; a22;]
% pole: Laguerre filter pole (pole=0 -> TDL; pole=0.7564 -> Bark)
%
% Output:
% Y filter output
% P Predicted output
% states final values of the delay line
%
% Author: Teun Selten
% Date: 17 September 2003
% Notes: - X(:,1) = x1 is Left
% - X(:,2) = x2 is Right
% - structure X = [ (x1[n-K] x1[n-K+1] x1[n-K+2] ... x1[n-1] x1[n])',
% (x2[n-K] x2[n-K+1] x2[n-K+2] ... x2[n-1] x2[n])' ]
% - Toeplitz can be one row smaller because this is the next state

function [S, HH] = stereo_spectra(apfile, lambda, freq);

%DESCRIPTION
% Analysis filter transfers for a 2-channel LP system
%INPUT
% apfile: prediction parameters (4 columns)
% lambda: Laguerre parameter
% freq frequency axis
% phi rotation angle
%OUTPUT
% S : matrix of length(freq)x2 containing the eigenvalues for each frequency
% HH: partial transfers of the 2-channel analysis system

function [alpha, y1, y2, depth, al] = rota_opt(x1, x2);

%DESCRIPTION
% Calculates the optimal rotation given two input signals
% and applies it to form a main and side signal
%INPUT
% x1 : first signal
% x2 : second signal
%OUTPUT
% alpha : optimal rotation angle
% y1 : first rotated signal (main component)
% y2 : second rotated signal (secondary component)
% depth : modulation depth (0 < md < 1)
%AUTHOR
% Bert den Brinker
%ASSUMPTIONS
% x1 and x2 have the same size (vectors)
% x1 and x2 are real-valued

function [g, s1, s2] = quant_sig_uni(sig, U, Levels);
%DESCRIPTION
% returns quantized signal and gains
%INPUT
% sig : signal
% U : segment size
% Levels: number of representation levels per sample
%OUTPUT
% g : quantization step per segment
% s1 : representation levels
% s2 : quantized signal

function out = nonuniformquant(sig, U, levels);
% Description:
% Quantize the input signal with a non uniform quantizer
% Input:
% sig: the to be quantized input signal
% U: block-length
% levels: number of levels (must be an odd number)
% Output:
% out: quantized output signal
% Author: Teun Selten
% Date: 16 January 2004
% Notes: ---

function [y1, y2] = rotate(x1, x2, alpha);
% Description:
% Rotates the given two input signals over alpha
% Input:
% x1 : first signal
% x2 : second signal
% alpha : rotation angle
% Output:
% y1 : first rotated signal
% y2 : second rotated signal
% Author: Teun Selten
% Date: 10 October 2003
% Notes:
% from function rota_opt from Bert den Brinker
% x1 and x2 have the same size (vectors)
% x1 and x2 are real-valued

MatLab call for Mex-file
out = dec_Stereo_laguerre_fast(residual, apfile, datalength, pole, order);
%
% Input:
% residual : (1:length(input),2)
% apfile : (1:(order * #segments) ,4)
% datalength : data length of a segment
% pole : laguerre pole
% order : order of the synthesis filter
%
% Output:
% out: (1:length(input),2)
Bibliography

[1] Bauer, D. and D. Seitzer: "Statistical properties of high quality stereo signals in the
  time domain." *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP-89)*,


  11, no. 6, pp 509 - 519.


  Preprint 6072.

  York, USA, September 2003. Preprint 5924

  68 - 75.


  Approach in Audio Coding." *112th AES Conv.*, Munich, Germany, May 2002. Preprint
  5553.


73


