WARPED LINEAR PREDICTION FOR IMPROVED PERCEPTUAL QUALITY IN THE SCELP LOW DELAY AUDIO CODEC (W-SCELP)

Hauke Krüger and Peter Vary

Institute of Communication Systems and Data Processing
RWTH Aachen University, D-52056 Aachen, Germany
krueger@ind.rwth-aachen.de, vary@ind.rwth-aachen.de

ABSTRACT

The SCELP (Spherical Code Excited Linear Prediction) audio codec, which has recently been proposed for low delay audio coding [5], is based on linear prediction (LP). It applies closed-loop vector quantization employing a spherical code which is based on the Apple Peeling code construction rule. Frequency warped signal processing is known to be beneficial especially in the context of wideband audio coding based on warped linear prediction (WLP).
In this contribution, WLP is incorporated into the SCELP low delay audio codec. The overall audio quality of the resulting W-SCELP codec benefits from an improved perceptual masking of the quantization noise. Compared with existing standardized audio codecs with an algorithmic delay below 10 ms, the W-SCELP codec at a data rate of 48 kbit/sec outperforms the ITU-T G.722 codec at a data rate of 56 kbit/sec in terms of the achievable audio quality.

1. INTRODUCTION

Most of the popular audio codecs, e.g. the Advanced Audio Codec (AAC), [1], are based on perceptual audio coding. In perceptual audio coding in general an audio signal is at first transformed by an analysis filter bank. The resulting representation in the transform domain is quantized whereas a perceptual model controls the adaptive bit allocation. Large transform lengths cause a high algorithmic delay. Considering mobile communications, the approach of linear predictive coding (LPC) has been followed for many years in speech coding. In LPC, an all-pole filter models the spectral envelope of an input signal. The signal is filtered with the inverse of that all-pole filter to produce the LP residual which is quantized. In the most recently standardized speech codecs, vector quantization (VQ) based on a sparse codebook is applied, following the CELP (Code Excited Linear Prediction) analysis-by-synthesis principle, [2]. A well-known example for this approach is the adaptive multi rate speech codec (AMR), [3]. Due to the sparseness of the codebook and modeling of the speakers instantaneous pitch period, speech coders can not compete with perceptual audio coding for non-speech input signals. The algorithmic delay is in general lower than that in perceptual coding.
The new SCELP audio codec targets application scenarios which require high audio quality and a very low algorithmic delay, for example digital audio transmission for a wireless headphone. It employs the principle of combined linear prediction and vector quantization (LP VQ) as known from speech coding. In order to achieve a better perceptual audio quality than speech coders, a spherical codebook is employed. The spherical codebook is constructed according to the Apple Peeling principle. This principle was introduced in [4] for the purpose of channel coding. In [5] we have proposed an efficient vector search procedure for the spherical codebook for linear predictive quantization, and in [6] a representation of the available Apple Peeling code vectors as coding trees has been introduced. Both techniques enable very efficient encoding and decoding with respect to computational complexity and memory consumption.
In [7] it was shown that warped signal processing techniques are suited to decrease the required data rate for wideband audio coding while retaining the same subjective audio quality. WLP is employed in a simulated coding system with D*PCM in that contribution. In contrast to that, in this contribution WLP will be incorporated into the closed-loop analysis-by-synthesis framework of the SCELP codec which was introduced for conventional LP primarily.
The principle of the SCELP audio codec and warped linear prediction will be introduced in Section 2 and 3 respectively. The modifications required for the application of WLP in analysis-by-synthesis VQ in general and the SCELP framework for highly efficient encoding in particular are described in Section 4. Results are presented in Section 5, including a comparison of the W-SCELP codec with the ITU-T G.722 [8] low delay audio codec.

2. PRINCIPLE OF THE SCELP AUDIO CODEC

The SCELP low delay audio codec is based on block adaptive combined linear prediction and vector quantization: The correlation immanent to an input signal $x(k)$ is exploited

in order to achieve a high quantization signal-to-noise-ratio (SNR). For this purpose, a windowed segment of the input signal of length L_{LP} is analyzed in order to obtain the N time-variant filter coefficients $a_1 \cdots a_N$. Based on these LP coefficients the LP analysis filter with system function $H_A(z) = 1 + \sum_{i=1}^{N} a_i \cdot z^{-i}$ converts the input signal into the LP residual signal $d(k)$ which is segmented into $N_V = L_{LP}/L_V \in \mathbb{N}$ non overlapping signal vectors $\mathbf{d} = \begin{bmatrix} d_0 & d_1 & \cdots & d_{L_V-1} \end{bmatrix}$ of length L_V. Each LP residual vector is quantized and transmitted to the decoder as code vector index i_Q. For signal reconstruction, also the LP coefficients must be transmitted to the decoder. In general this can be realized with only small additional bit rate as shown for example in [9]. In the decoder, the transmitted code vector index i_Q is the basis for the reconstruction of the quantized LP residual vector $\tilde{\mathbf{d}}$ which is filtered by the LP synthesis filter $H_S(z) = (H_A(z))^{-1}$. The output of the LP synthesis filter is the decoded signal vector $\tilde{\mathbf{x}}$ and hence the signal $\tilde{x}(k)$ [10].
The principle of the encoder of a CELP codec is depicted in Figure 1. The decoder is part of the encoder. According to

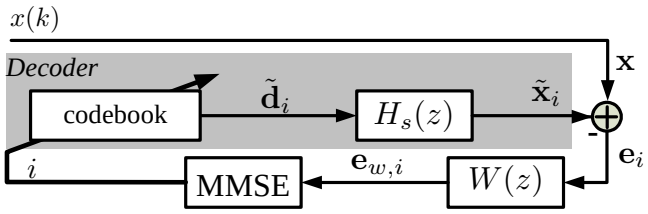


Figure 1: Scheme SCELP Audio Codec, Encoder.

the analysis-by-synthesis principle, the LP residual vector $\tilde{\mathbf{d}}_i$ for each codebook index i is generated first. This excitation vector is filtered by the LP synthesis filter $H_S(z)$ to obtain the corresponding decoded signal vector candidates $\tilde{\mathbf{x}}_i$. The error distance between the input signal and the decoded signal, $\mathbf{e}_i = \mathbf{x} - \tilde{\mathbf{x}}_i$, is determined for each vector candidate corresponding to index i. The goal is to find the index i_Q for which the minimum mean square error is achieved:

$$i_Q = \arg\min_i \{\mathcal{D}_i = \| \mathbf{e}_i \|^2 = (\mathbf{x} - \tilde{\mathbf{x}}_i) \cdot (\mathbf{x} - \tilde{\mathbf{x}}_i)^T\}. \quad (1)$$

The error weighting filter $W(z)$ controls the spectral shape of the quantization noise inherent to the decoded signal for perceptual masking of the quantization noise. The analysis-by-synthesis vector search can be exhaustive for a large vector codebook.

2.1. Spherical Vector Codebook

In the SCELP audio codec, vector quantization is applied in a *gain-shape* approach to encode the LP residual. Each LP residual vector $\mathbf{d}$ is decomposed into a radius for the *gain* and a vector on the surface of a unit sphere for the *shape* component. While the radius R is quantized by means of logarithmic scalar quantization, the valid code vectors for the quantization of the shape component are based on the Apple Peeling code construction rule. This rule was described and demonstrated for the special case of a 3-dimensional sphere in [5]. The design target of the Apple Peeling code is to place all codebook vectors on the surface of a unit sphere as uniformly as possible.
The **decoder** in CELP coding in general is not very complex. For a low computational **encoding** complexity, the analysis-by-synthesis approach in Figure 1 was modified in the SCELP encoder as described in [5]. The result is a low complexity vector search framework. Additionally, the technologies called *Pre-Selection* and *Candidate-Exclusion*, combined with an efficient metric computation, enable a very efficient code vector search. Furthermore the representation of the Apple Peeling code vectors as coding trees was explained in [6] for the sake of highly efficient encoding and decoding.

3. WARPED LINEAR PREDICTION

The principle and properties of warped linear prediction are discussed in [7]. In this contribution only those aspects that are relevant for the analysis-by-synthesis vector search of the SCELP will be briefly presented.
In conventional linear prediction the approximation of the spectral envelope of a signal is based on a uniform resolution of the frequency scale. Considering the perceptual properties of human hearing, a uniform resolution is known to be inferior compared to a non-uniform resolution of the frequency scale. For this purpose, a non-uniform resolution of the frequency scale is achieved by applying WLP. Considering the z-transform of a signal, this can be realized by replacing all unit delay elements by an allpass filter $AP(z)$,

$$z^{-1} \rightarrow AP(z) = \frac{z^{-1} - \lambda}{1 - \lambda \cdot z^{-1}} \quad | \lambda | < 1; \lambda \in \mathbb{R} \quad (2)$$

For positive values of warping constant λ, the spectral resolution is increased for lower and decreased for higher frequencies compared to conventional LP.

3.1. Warped LP Analysis

In the SCELP codec the LP analysis is based on the auto correlation method, as for example described in [10]. In [7], it was shown that for the warped LP analysis, in the auto correlation method all unit delay elements must be replaced by the first order allpass filter $AP(z)$ according to (2). Hence the warped auto correlation coefficients $\varphi^w_{x,x}(0) \cdots \varphi^w_{x,x}(N)$ are determined as demonstrated for the first three coefficients in Figure 2. Warped auto correlation coefficients can

be transformed into warped LP coefficients $a_1^w \cdots a_N^w$ by means of the Levinson Durbin algorithm as in conventional LP.

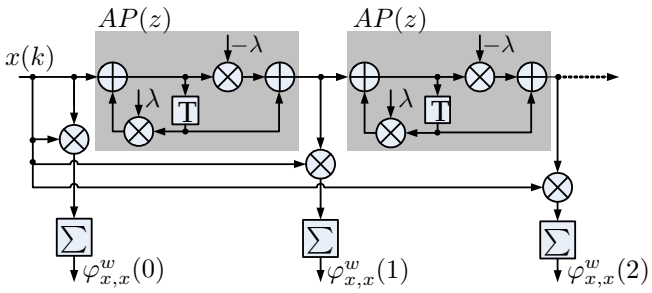


Figure 2: Warped LP Analysis.

3.2. LP Analysis/Synthesis filter

For the warped LP analysis and synthesis filter, all unit delay elements of the conventional LP analysis/synthesis filter are replaced by allpass filters $AP(z)$:

$$H_A^w(z) = H_A^w(AP(z)) = 1 + \sum_{i=1}^{N} a_i^w \cdot AP(z)^i = (H_S^w(z))^{-1}. \tag{3}$$

The filter coefficients a_i^w are calculated according to Section 3.1.

3.3. Error Weighting Filter

The SCELP audio codec employs an error weighting filter as proposed in [11]. In conventional linear prediction this error weighting filter can be calculated from the LP analysis filter:

$$W(z) = \frac{H_A(z/\gamma_2)}{H_A(z/\gamma_1)}. \tag{4}$$

The coefficients γ_1 and γ_2 are within the range of $0 \leq \gamma_1 \leq \gamma_2 \leq 1.0$ and control the degree of noise shaping. With the application of the error weighting filter, the quantization noise inherent to the decoded output signal is spectrally shaped according to the system function of the inverse of the error weighting filter, $(W(z))^{-1}$.
Considering WLP, all unit delay elements in equation (4) must be replaced by $AP(z)$ in the warped error weighting filter:

$$W^w(z) = \frac{H_A^w(AP(z) \cdot \gamma_2)}{H_A^w(AP(z) \cdot \gamma_1)}. \tag{5}$$

4. WLP IN THE SCELP CODEC

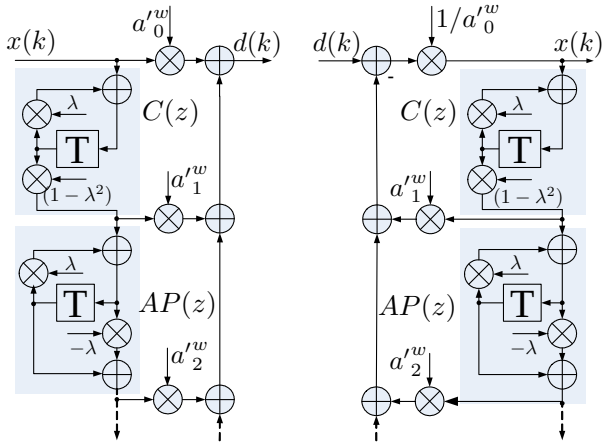
The properties of the warped linear prediction prohibit a straight forward incorporation into the SCELP audio codec. Therefore the following modifications must be considered first.

4.1. Zero-Delay Path in Feedback Loop

A zero-delay path in the feedback loop makes the implementation of the LP synthesis filter according to equation (3) impractical. In contrast to the implementation of the warped LP synthesis filter in [7], in this contribution [12] the substitution of

$$C(z) = AP(z) + \lambda \tag{6}$$

is applied to the first allpass filter in the allpass chain of the warped LP synthesis filter to remove the zero-delay path. The resulting filter structure is employed for warped LP analysis and synthesis filter as depicted in Figure 3, and also for the error weighting filter (5).



(a) WLP Analysis Filter (b) WLP Synthesis Filter

Figure 3: Modified Structure for Warped LP Filters.

As a consequence of the applied substitution, modified filter coefficients $a'^w_0 \cdots a'^w_N$ are used in the new LP analysis and synthesis filter structure. These can be calculated from the original coefficients $a_0^w \cdots a_N^w$ recursively as

$$\begin{aligned} a'^w_N &= a_N^w \\ a'^w_i &= a_i^w - \lambda \cdot a'^w_{i+1}; \quad i = N-1, \cdots 0; \; a_0^w = 1.0 \end{aligned} \tag{7}$$

4.2. Zero-Mean Property

Decorrelation of an input signal without any additional amplification is connected to the well-known zero-mean property in conventional LP [13]. WLP does not provide this

property. It can be shown, however, that the WLP filters have zero-mean property if the modified filter coefficients ${a'}_i^w$ (7) are normalized according to

$$ {a''}_i^w = {a'}_i^w / {a'}_0^w \qquad i = N, \cdots, 0. \tag{8}$$

Considering Figure 3, the filter coefficients ${a'}_i^w$ must be replaced by the normalized coefficients ${a''}_i^w$, with the first coefficient resulting to ${a''}_0^w = 1.0$.

4.3. Spectral Tilt Compensation

Due to the non uniform resolution of the frequency scale in warped signal processing, the warped LP residual signal is not perfectly flat but contains a spectral tilt [14]. This spectral tilt inherent to the LP residual must be compensated prior to quantization to achieve the highest quantization SNR in closed-loop LP VQ. For this purpose the filters according to (3), in the structure as depicted in Figure 3, employing the normalized coefficients ${a''}_i^w$ (8), are operated in the cascade with the tilt compensation filter

$$ H_t^w(z) = 1 - \lambda \cdot z^{-1} \tag{9}$$

to form the overall LP analysis/synthesis filter [1]

$$ H_A^{w,t}(z) = H_A^w(z) \cdot H_t^w(z) = (H_S^{w,t}(z))^{-1}. \tag{10}$$

4.4. SCELP Low Complexity Vector Search

Considering the analysis-by-synthesis principle, a low complexity vector search procedure is employed in the SCELP codec. It enables to search the large spherical vector codebook in a very efficient way to achieve the low computational complexity of the codec. The principle was introduced in [5] and is depicted in Figure 4.

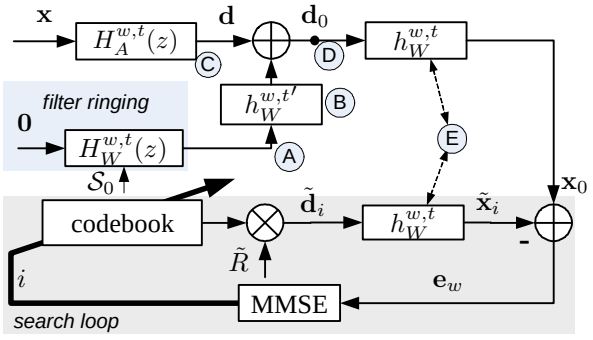


Figure 4: Modified Analysis-by-synthesis.

[1] No tilt compensation is required for the error weighting filter because the same warping factor is used for numerator and denominator part.

Considering Figure 4, those functional blocks which must be adapted for the application of WLP will be identified in the following:
The basis for an efficient vector search in the SCELP is the determination of signal $\mathbf{d}_0$ prior to the actual analysis-by-synthesis vector search procedure. In order to determine this signal, the filter ringing signal, marked by the label A, must be obtained first. Latter is the filter output resulting from the history related to previously quantized signal frames. Before a new signal input vector $\mathbf{x}$ is quantized, this history is stored as the filter states $\mathcal{S}_0$. In order to get the vector related to the filter ringing, the filter $H_W^{w,t}(z)$ is fed with a zero input signal vector $\mathbf{0}$ of length L_V. For the W-SCELP, this filter is identified as the cascade of the warped LP synthesis filter according to (10) and the error weighting filter according to (5):

$$ H_W^{w,t}(z) = H_S^{w,t}(z) \cdot W^w(z). \tag{11}$$

In order to obtain signal vector $\mathbf{d}_0$, the filter ringing signal vector must be transformed into the residual signal domain. This is done in Figure 4 by means of convolution, marked as the block $h_W^{w,t'}$ at position B. $h_W^{w,t'}$ is identified as the truncated impulse response of the inverse of filter $H_W^{w,t}(z)$ (11):

$$ h_W^{w,t'} = \left[{h_{W,0}^{w,t}}' \cdots {h_{W,(L_V-1)}^{w,t}}' \right]; \quad {h_{W,k}^{w,t}}' \circ\!\!-\!\!\bullet\, (H_W^{w,t}(z))^{-1} \tag{12}$$

The convolved filter ringing is added to the warped LP residual $\mathbf{d}$. The LP residual $\mathbf{d}$ is obtained by filtering $\mathbf{x}$ in the warped LP analysis filter $H_A^{w,t}(z)$, position C. The resulting signal vector $\mathbf{d}_0$, position D, is analyzed to determine the corresponding radius R which is quantized as $\tilde{R}$.
In the analysis-by-synthesis vector search procedure, code vectors $\tilde{\mathbf{d}}_i$ are generated by multiplying the spherical shape-component vector candidates with the quantized radius $\tilde{R}$. Considering the metric (1) to find the optimal excitation vector $\tilde{\mathbf{d}}_{i_Q}$, the unquantized and the quantized residual vector candidate, $\mathbf{d}$ and $\tilde{\mathbf{d}}_i$, both must be transformed from the LP residual into the signal domain. For this purpose, the two blocks $h_W^{w,t}$, position E, represent the transform for both signals by means of convolution with the truncated impulse response of filter $H_W^{w,t}(z)$ (11):

$$ h_W^{w,t} = \left[h_{W,0}^{w,t} \cdots h_{W,(L_V-1)}^{w,t} \right]; \quad h_{W,k}^{w,t} \circ\!\!-\!\!\bullet\, H_W^{w,t}(z) \tag{13}$$

Now that all functional blocks, which were introduced for the SCELP codec, have been identified also for the W-SCELP codec, the principles of *Pre-Selection*, the efficient metric computation and the *Candidate-Exclusion* explained in [5] for highly efficient encoding can be applied also in the W-SCELP.
With the determination of quantized LP residual vector $\tilde{\mathbf{d}}_{i_Q}$,

the differential signal $\mathbf{d} - \tilde{\mathbf{d}}_{i_Q}$ must be processed by filter $H_W^w(z)$ to finally determine the update for the filter states $\mathcal{S}_0$ restored for the quantization of the next signal frame.

4.5. Complexity

The computational complexity of the warped LP analysis, synthesis and error weighting filter in the W-SCELP codec is higher than that of the same filters realized for conventional LP in the SCELP codec. The biggest part of the overall complexity of the SCELP codec, however, is spent on the analysis-by-synthesis vector search. Since the W-SCELP benefits from the same principles targeting low complexity encoding as the SCELP, the overall complexity is only marginally increased. The complexity of the encoder of the conventional SCELP codec was estimated as 20-25 WMOPS in [5], that of the encoder of the W-SCELP codec as 23-28 WMOPS. The decoder of the W-SCELP codec has an estimated complexity of 2-3 WMOPS.

5. RESULTS

For the comparison of the achieved quality of the W-SCELP and the SCELP codec, both codecs have been configured identically for a sample rate of $f_s = 16$ kHz. The resulting overall data rate is approximately 48 kbit/sec, and the noise shaping coefficients have been set to $\gamma_1 = 0.6$ and $\gamma_2 = 0.94$. The order of the linear prediction in both cases is $N = 10$ and the algorithmic delay $L_{LP}\hat{=}9$ ms. For the W-SCELP codec, the highest performance has been determined for a warping factor $\lambda = 0.46$ in informal listening tests.

Comparing the prediction gain in W-SCELP and SCELP as a measure of signal decorrelation, WLP provides only an insignificantly higher value. Considering perceptual masking of the quantization noise, it was observed in informal listening tests that the higher spectral resolution of WLP for lower frequencies provides significant benefits. Especially for audio signals with a sparse spectrum, for example the sound of a flute, WLP provides clearly better perceptual results than conventional LP.

Considering a formal assessment of the quality, speech was processed by the W-SCELP and the SCELP codec. The decoder output was rated with the WB-PESQ measure [15] which is widely used in the speech coding community. As result, the W-SCELP outperformed the SCELP by 0.2 on the MOS scale. Comparable results may also be obtained using the PEAQ quality measure [16].

For a comparison of the W-SCELP codec with a standardized audio codec, the same speech signal was also processed by the ITU-T G.722 low delay audio codec at 48, 56 and 64 kbit/sec. This reference codec was chosen because of its algorithmic delay in the magnitude of that of the W-SCELP codec (below 10 ms)[2]. The result of the formal comparison of the new codec with the G.722 reference codec is listed in Table 1 in the order of descending perceptual quality. The

Codec	G.722 mode 1	**W-SCELP**	G.722 mode 2	G.722 mode 3
Data rate	$64\frac{\text{kBit}}{\text{sec}}$	$48\frac{\text{kBit}}{\text{sec}}$	$56\frac{\text{kBit}}{\text{sec}}$	$48\frac{\text{kBit}}{\text{sec}}$
WB-PESQ (MOS-LQO)	4.47	4.4	4.39	4.02

Table 1: Results Formal Quality Assessment.

performance of the G.722 codecs was rated with 4.02, 4.39 and 4.47 MOS for the three codec modes respectively. The W-SCELP at a data rate of roughly 48 kbit/sec reached a value of 4.4 MOS. Considering this result, the quality of the W-SCELP codec at 48 kbit/sec can be classified as slightly better than that of the G.722 at 56 kbit/sec.

6. CONCLUSION

In this contribution the principle of warped signal processing was incorporated into the new SCELP low delay audio codec to form the W-SCELP codec. While the overall complexity of the W-SCELP is only insignificantly higher than that of the SCELP codec, the achievable audio quality is clearly better. In a comparison with a standardized codec that has a similar algorithmic delay, the W-SCELP at a data rate of 48 kbit/sec outperforms the ITU-T G.722 audio codec at a data rate of 56 kbit/sec.

7. REFERENCES

[1] ISO/IEC 13818-7, "Advanced Audio Coding (AAC)," 1997.

[2] M. Schroeder and B. Atal, "Code-excited Linear Prediction (CELP): High-quality Speech at very low Bit Rates," *Proc. ICASSP*, 1985.

[3] Rec. GSM 06.90 ETSI, "Adaptive Multi-Rate (AMR) Speech Transcoding," 1998.

[4] E. Gamal, L. Hemachandra, I. Spherling, and V. Wei, "Using Simulated Annealing to Design Good Codes," *IEEE Trans. Inform. Theory*, vol. it-33, 1987.

[5] H. Krüger and P. Vary, "SCELP: Low Delay Audio Coding with Noise Shaping based on Spherical Vector Quantization," *EUSIPCO, Florence, Italy*, 2006.

[2] An alternative codec with a comparable algorithmic delay is the ULD codec [17]. This codec, however, is not freely available and was thus not considered.

[6] H. Krüger and P. Vary, "An Efficient Codebook for the SCELP Low Delay Audio Codec," *MMSP, Victoria, Canada*, 2006.

[7] A. Härmä and U. Laine, "A Comparison of Warped and Conventional Linear Predictive Coding," *IEEE Trans. Speech and Audio Processing*, vol. 9, no.5, 2001.

[8] ITU-T Rec. G.722, "7 kHz Audio Coding within 64 kbit/s," 1988.

[9] K. Paliwal and B. Atal, "Efficient Vector Quantization of LPC Parameters at 24 Bits/Frame," *IEEE Trans. Speech and Signal Proc.*, vol. 1, no.1, pp. 3–13, 1993.

[10] N. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice-Hall, Inc., 1984.

[11] M. Schroeder, B. Atal, and J. Hall, "Optimizing Digital Speech Coders by Exploiting Masking Properties of the Human Ear," *Journal of the Acoustical Society of America*, pp. 1647–1652, 1979.

[12] K. Steiglitz, "A Note on Variable Recursive Digital Filters," *IEEE Trans. Acc., Speech, and Signal Processing*, vol. 28, 1980.

[13] J. Markel and A.Gray, *Linear Prediction of Speech*, Springer, 1976.

[14] H. W. Strube, "Linear Prediction on a Warped Frequency Scale," *Journal of the Acoustical Society of America*, vol. 68, pp. 1071–1076, 1980.

[15] ITU-T Rec. P.862.2, "Wideband Extension to Recommendation P.862 for the Assessment of Wideband Telephone Networks and Speech Codecs," 2005.

[16] Recommendation ITU-R BS.1387, "Method for objective measurements of perceived audio quality," 2001.

[17] http://www.idmt.fraunhofer.de, "Audio Coding with Ultra Low Encoding/Decoding Delay," 2007.