WARPED LINEAR PREDICTIVE AUDIO CODING IN VIDEO CONFERENCING APPLICATION

Kalle Palomäki, Aki Härmä and Unto K. Laine
Helsinki University of Technology
Laboratory of Acoustics and Audio Signal Processing
P. O. Box 3000, 02015, Espoo, Finland
e-mail: kpalomak@cc.hut.fi

ABSTRACT

A codec for wideband 12kHz speech and audio in a video conferencing application is proposed in this paper. The codec is based on warped linear predictive coding algorithm which utilizes the auditory Bark frequency resolution. The structure of the codec is described and the main issues on the real-time implementation are discussed. The codec is integrated to a video conferencing product which is a video codec PC-board. The algorithm is implemented in a Texas TMSC31 digital signal processor.

1 Introduction

In video conferencing applications the transmission of high quality speech and audio is a very important task. In such a system the main part of the transmission bandwidth is used for coded video data, increasing video data bit-rate provides higher quality picture. The aim of audio coding in a video conferencing systems is to provide good quality audio at as low bit-rate as possible, so that more transmission capacity is left for video data. Nowadays the video conferencing standard H.320 [1] includes three standards for audio coding: two voiceband codecs (300Hz-3.5 kHz) [2], [3], and a wide band speech codec (0-7kHz) [4]. Voiceband codec provides transmission of understandable speech. Higher quality of speech and audio is achieved using the 7kHz audio algorithms. However, there is still need to enhance sound quality and increase audio bandwidth. In this paper a computationally efficient and low complexity wide band audio coding algorithm is applied to a video conferencing system.

The linear prediction has been traditional method in speech coding, since it is an effective source model of speech organ. In high quality speech or audio coding it is also important to consider to the properties of the hearing not only the speech organ. The uniform frequency resolution of linear prediction differs greatly from the nonuniform frequency resolution of the hearing. The algorithm applied in this work utilizes warped linear prediction (WLP), where frequency resolution approximates auditory frequency resolution (Bark scale) [5]. WLP was first introduced in speech coding application by Strube [6]. Since then WLP has been applied in the coding of speech and wideband audio in [5, 7, 8, 9]. The WLP-codec presented in this article is based on the basic and simple version of WLP, introduced by Härmä [7].

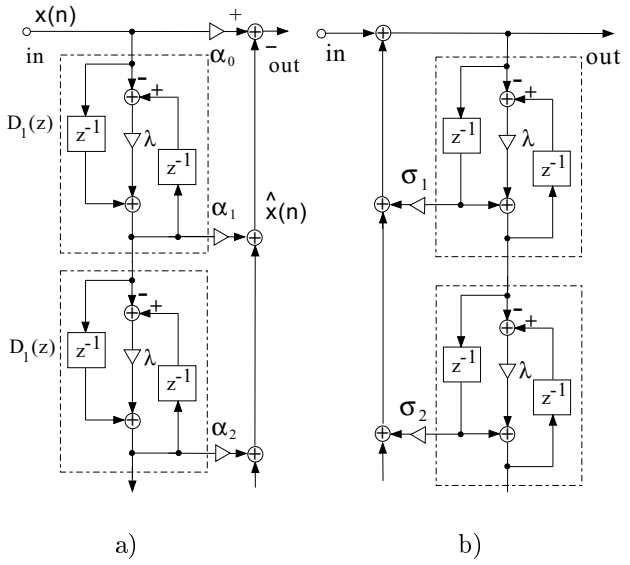

Figure 1: *a) Inverse filter (WFIR) of WLP-codec. b) Synthesis filter (WIIR) of WLP-codec.*

2 Warped Linear Prediction (WLP)

Here the derivation of WLP is started from an ordinary LP. In an M:th order linear predictor the current sample is predicted as linear combination from M previous samples. The predictor given as

$$\hat{x}(n) = \sum_{j=1}^{M} \alpha_j x(n-j) \qquad (1)$$

forms a FIR filter. The predictor coefficients i.e. α-coefficients can be estimated from a signal $x(n)$ using the autocorrelation method [10].

In warped linear prediction an estimate for a sample is not produced from the previous values as in the case of

conventional LP, but the samples of a frequency warped signal. The warped predictor is derived replacing the unit delays in the predictor of conventional LP with allpass elements which acts as frequency dependent unit delays. Transfer function of a warped predictor is

$$H(z) = \sum_{n=0}^{M} \alpha_n D_1(z)^n \tag{2}$$

where $D_1(z)$ is an allpass element. Transfer function of $D_1(z)$ is given by

$$D_1(z) = \frac{z^{-1} - \lambda}{1 - \lambda z^{-1}} \tag{3}$$

Figure 1a shows WFIR type inverse filter, where part that outputs $\hat{x}$ is the warped predictor. α-coefficients for a warped predictor are estimated using the warped autocorrelation function instead of autocorrelation of conventional LP.

However the implementation of recursive warped filter structures such as WIIR-filters is more problematic than just replacing the unit delays with allpass elements. In such structures direct replacement leads to delayless loops which are not straigthforward to implement. In the current WLP-codec the recursive structure is needed in the synthesis filter of the decoder. There exist several solutions for the realization of WIIR-structures. In this case it is modified so that the delayless loops are avoided [11]. The synthesis filter containig the modified structure is shown in figure 1b. With this structure the α-coefficients of WFIR filter are not valid and they has to be mapped to σ-coefficients of the modified structure. The mapping function is descriped in [12]. The recursive structures can also be implemented without using any modifications in the filter structure. Härmä [13] showed a general solution for the implementation of delayless loops.

3 Structure of WLP-codec

The structure of WLP-codec is presented in figure 2. The structure of the encoder is a prediction error coder [14], where the quantization process is fed with the prediction error signal $\hat{x}$. The encoder can be separated functionally in three separate tasks: 1) pre-emphasis filtering, 2) warped coefficient estimation and coefficient quantization, 3) inverse filtering and residual, i. e., prediction error signal quantization.

Before the coefficient estimation the input signal is filtered with a pre-emphasis filter. Usually an audio or speech signal has more energy in the lower frequencies than in the higher frequencies. The WLP- or LP estimation process models more effectively the input spectrum, where the energy is distributed more uniformly over the whole frequency range. The tilt of the spectrum is usually corrected with pre-emphasise filtering of the input. Pre-emphasis filter is a first order high pass filter which

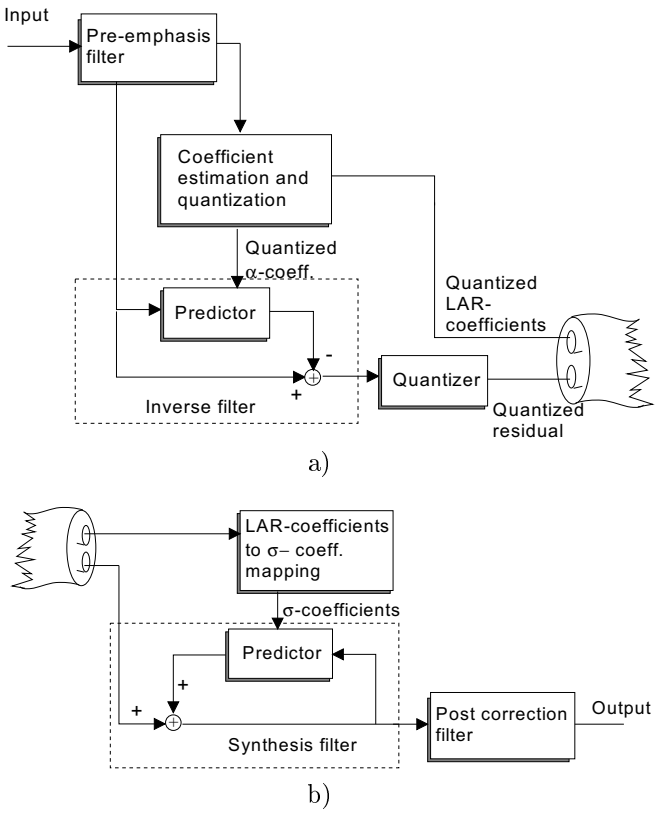

Figure 2: *Block diagram of WLP-codec. a) Encoder b) Decoder*

should approximate the inverse of the spectral tilt of the input signal.

The warped autocorrelation coefficients are estimated with the warped autocorrelation function. Before coefficient estimation process a signal is windowed with rectangular overlapping window, since it provides better inverse filter responses especially on low frequencies. From the set of warped autocorrelation coefficients, reflection or α- coefficients can be solved using Levinson-Durbin recursion. In this case reflection coefficients are mapped to log area ratios (LAR) which are then quantized for transmission. LAR-representation for the filter coefficients is conventionally used in speech codecs, since they are more suitable for quantization than the α-coefficients [10]. Quantized LAR-coefficients are mapped to α-coefficients for the inverse filter, because the same set of quantized coefficient has to be both transmitted to decoder and used in the inverse filter.

The output of the inverse filter is prediction error i.e. residual signal, which is a difference between signal value and the predictor value. The residual is quantized using simple 2-bit backward adaptive one memory word quantizer [15].

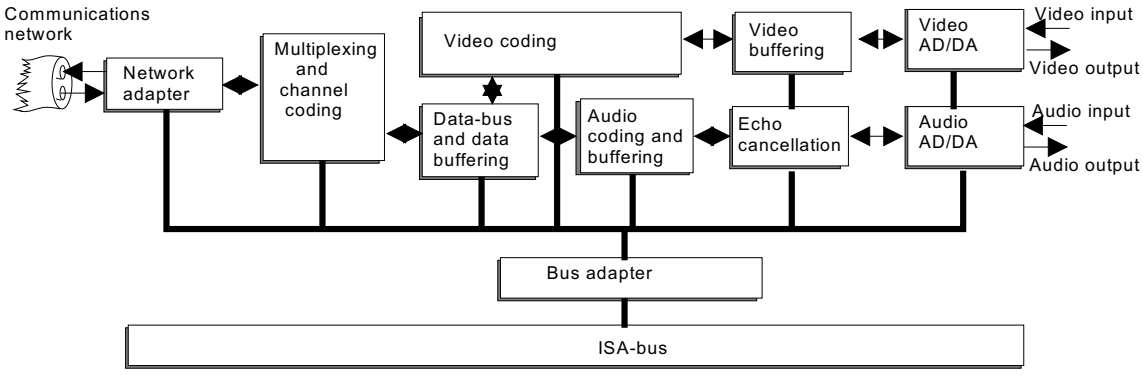

Figure 3: *Block diagram of video conferencing codec of VistaCom*

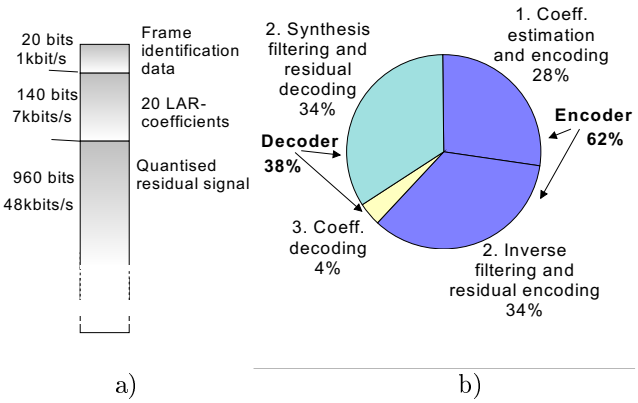

Figure 4: *a) The structure of WLP-transmission frame. b) The division of computational time between different parts of coding algorithm.*

For the synthesis filter the LAR-coefficients are mapped to WIIR σ-coefficients instead of α-coefficients of inverse filter. The output singal of synthesis filter involves still the spectral shape of pre-emphasis filter. The original shape is returned with a post correction filter which is an inverse of pre-emphasis.

4 The real-time implementation

WLP-codec is implemented to a video codec PC-board of Vistacom Inc. [16]. It is a part of video conferencing system, where communication frame packing, real-time coding video and audio is performed. A simplified block diagram of the board is shown in figure 3. Video codec board is designed for the implementation of the video conferencing standard H.320 [1]. H.320 defines necessary standardization for the compatibility between video conferencing products of different manufacturers. The H.320 includes standardization in such areas as video coding, audio coding and channel coding. The audio coding standards and the WLP-codec are implemented in a TMS320C31 floating point digital signal processor.

Standard/algorithm	*Transmission bw.*	*Audio bw.*
G.711A,G.711μ	48-64 kbits/s	3.4 kHz
G.728	16 kbits/s	3.4 kHz
G.722	48-56 kbits/s	7 kHz
WLP	56 kbits/s	12 kHz

Table 1: *Transmission and audio bandwidths of audio coding algorithms and WLP-codec*

The performance of 60 MHz 'C31 is 60 MFLOPS and 30 MIPS. Sample rate with the current version of the WLP-codec is 24kHz i.e. the audio bandwidth is 12kHz. The aim of design of WLP-codec was to increase audio bandwidth from 7.5kHz of G.722 upto 11-12 kHz with the same 56kbit/s transmission band as G.722 codec. Table 1 shows a comparation of transmission and audio bandwidths of standard codecs and WLP-codec.

The high performance requirement of the WLP-codec with the such high sample rate as 24kHz led to compromises between transmission bandwidth and the computational load of the codec. Effective vector quantization techniques seemed to be impossible to use in this application. Jayant's adaptive one memory word quantizer seemed to provide computationally low cost solution for the residual and the coefficient quantization [15]. Figure 4 shows the structure of the whole transmission frame. It starts with the frame identification data. The next part is 20 LAR coefficients where 7bits are used for each. The largest part of the frame consist of 2bit/sample quantized residual signal. The length of audio frame in time is 20 ms. With 24 kHz sample rate the transmission band of residual becomes 2bits/sample*24kHz = 48 kbits/s, where the whole bandwidth is 56 kbits/s. The remaining 8kbits/s is left for coefficients and additional information.

Typically in audio and speech coding schemes the computational complexity of encoder is much higher compared to decoder. With the current WLP-codec the difference between them is not so dramatical as for ex-

ample in MPEG 1 layer III codec, where encoder consumes significantly more processing time than the decoder. Figure 4 shows the approximative division of computation time between different parts of the WLP-algorithm. The division between encoder and decoder is 62% and 38%, where the most time consuming parts seem to be inverse filter with residual encoding and the synthesis filter with decoding of residual (both 34%). Close to them is coefficient estimation and quantization with 28%.

The H.320 standard does not specify the network connection. Usual connection types are the ISDN, and several local area network configurations. With such connections probability of bit errors in transmission may be assumed low. However transmission errors such as bit errors or missing parts of encoded data has to be prepared somehow. This fact sets limits for the adaptations over frame borders. A typical solution is to design codec so that each individual frame is independent from others and there are no adaptations over frame borders. However in this case those adapations have importance in coefficient quantization. So the adaptations over frame borders are limited over a finite set of frames. This partly adaptive technique enhances the behavior of coefficient quantization comparing to entirely fixed value quantizer.

5 Discussion

The performance limits within 'C31 environment led to compromises between the complexity of algorithm and transmission bandwidth. The most critical part of optimization work were the implementation warped filters and warped autocorrelation function. The design of LAR-coefficient quantizers is rather simple, where 7bits for each coefficients are used. With modern vector quantization methods it is possible to reach such low values as 2-3 bits/coefficient. In this case however, the amount of coefficient data is still only a small part of whole transmission frame. The line rate limit of 56kbits/s offered enough space for 7bit coefficients. More important point was computationally low cost design.

Always an interesting question with LPC-based speech codecs is the quality of non-speech audio. The ability to provide acceptable quality with various audio material and high quality speech led to a compromise also in the this case. Important "tuning parameters" having differences between speech and non-speech audio seemed to be adaptation parameters of residual quantizers. Optimal selection of quantizer parameters are discussed in [15].

6 Acknowledgement

The implementation work has been done in VistaCom Inc. The authors want to thank VistaCom people for the support and specially Mr. Esko Aalto and Mr. Tuomas Koljonen for the support and coordination of the project. The author are also grateful to Mr. Pertti Ylönen and Mr. Harry Santamäki, who activated the project.

References

[1] ITU-T Recommendation H.320, "Narrow-band visual telephone systems and terminal equipment," 1997.

[2] ITU-T Recommendation G.711, "Pulse code modulation (pcm) of voice frequencies," 1988.

[3] ITU-T Recommendation G.728, "Coding of speech at 16 kbit/s using low-delay code exited linear prediction," 1992.

[4] ITU-T Recommendation G.722, "7 khz audio coding within 64 kbits/s," 1988.

[5] U. K. Laine and M. Karjalainen, "WLP in speech and audio processing," in *ICASSP*, vol. III, (Adelaide), pp. 349–352, 1994.

[6] H. W. Strube, "Linear prediction on a warped frequency scale," *J. of the Acoust. Soc. Am.*, vol. 68, no. 4, pp. 1071–1076, 1980.

[7] A. Härmä, U. K. Laine, and M. Karjalainen, "WLPAC – a perceptual audio codec in a nutshell," in *AES 102nd Conv. preprint 4420*, (Munich), 1997.

[8] A. Härmä, U. K. Laine, and M. Karjalainen, "An experimental audio codec based on warped linear prediction of complex valued signals," in *ICASSP 97*, vol. 1, (Munich), pp. 323–327, 1997.

[9] K. Koishida, K. Tokuda, T. Kobayashi, and S. Imai, "Celp coding system based on mel-generalized cepstral analysis," in *Proc. of ICSLP'96*, vol. 1, 1996.

[10] J. D. Markel and A. H. Gray, *Linear Prediction of Speech*. Spirnger-Verlag, Berlin, Germany, 1976.

[11] T. Kobayashi, S. Imai, and Y. Fukuda, "Mel-generalized log spectral approximation filter," *Trans. IECE*, vol. 68, pp. 610–611, 1985.

[12] M. Karjalainen, A. Härmä, and U. K. Laine, "Realizable warped IIR filters and their properties," in *ICASSP'97*, vol. 3, (Munich), pp. 2205–2209, 1997.

[13] A. Härmä, "Implementation of recursive filters having delay free loops," in *Submitted to ICASSP'98*, (Seattle), 1998.

[14] N. S. Jayant and P. Noll, *Digital coding of waveforms*. Prentice-Hall inc., 1984.

[15] N. S. Jayant, "Adaptive quantization with one-word memory," *Bell Syst. Tech. J.*, pp. 1119–1144, 1973.

[16] VistaCom Inc. `<URL:http://www.vistacom.fi>`.