WIDEBAND SPEECH CODING USING FORWARD/BACKWARD ADAPTIVE PREDICTION WITH MIXED TIME/FREQUENCY DOMAIN EXCITATION

Jürgen Schnitzler, Joachim Eggers, Christoph Erdmann, Peter Vary*

Institute of Communication Systems and Data Processing (IND)
Aachen University of Technology (RWTH), D-52056 Aachen, Germany

Phone: +49 241 80-6982; Fax: +49 241 8888-186
E-mail: {Juergen.Schnitzler, Christoph.Erdmann, Peter.Vary} @ind.rwth-aachen.de

ABSTRACT

This paper describes a wideband (7 kHz) speech coding scheme using code-excited linear prediction (CELP) with mixed time and frequency domain excitation. The proposed frequency domain innovation can be used alternatively or in parallel to a time domain codebook. In addition an improved synthesis filter is used consisting of a signal dependent combination of a forward adaptive and a backward adaptive (FA/BA) structure. An experimental codec operating at 15.5 or 20.0 kbit/s is demonstrated.

1. INTRODUCTION

In wideband (7 kHz) speech coding linear predictive analysis by synthesis (AbS) techniques including long term prediction (LTP) are widely used at bit rates of 1-1.5 bits per sample (16-24 kbit/s). In [1, 2] we have combined subband (SB) coding and CELP techniques and achieved high speech quality at 16 kbit/s and below.
However, as non-speech signals such as music are attaining an increasing attention, further improvements are necessary. In [3] most of the ITU-T Q.20/16 requirements for speech and music at 16 and 24 kbit/s could be fulfilled by a scheme called ATCELP which switches between SB-CELP and adaptive transform coding (ATC), depending on the characteristics of the input signal.
Two other key strategies for improving the coding performance are

- to consider a frequency domain excitation technique such as TCX (*transform coded excitation*) [4], TPC (*transform predictive coding*) [5], or [6];
- to increase the spectral resolution of the synthesis filter by switched or combined forward/backward adaptive linear prediction (FA/BA-LP) [7, 8, 9].

In this paper, we propose a novel TCX approach in section 2. In section 3, we extend the combined FA/BA-LP coding approach of [9]: the usage of the BA-LP filter now depends on the instantaneous input signal characteristics and the expected codec performance. An experimental AbS codec scheme combining these elements is described in section 4, operating at bit rates of 15.5 and 20 kbit/s.

* Now with University of Erlangen-Nürnberg (LNT)
E-mail: eggers@nt.e-technik.uni-erlangen.de

2. FREQUENCY DOMAIN EXCITATION CODEBOOKS

In this proposed TCX codec, the input signal s is subject to a conventional LP analysis. The inverse LP filter, described by the transfer function $A(z)$, is used to obtain the residual vector $\mathbf{d}_i$ of length N from the input signal vector $\mathbf{s}_i$ in subframe i. Furthermore, a closed-loop LTP analysis yields the optimum adaptive codebook (ACB) vector $\mathbf{c}_{a,i}$.
The remaining encoder operation depicted in Fig. 1 consists of the AbS coding of the innovation, i.e. the difference $\mathbf{d}_i - \mathbf{c}_{a,i}$. In a conventional time domain CELP, synthesis filtering and perceptual spectral weighting of $\mathbf{d}_i - \mathbf{c}_{a,i}$, taking the ringing of the previous block into consideration, would result in a target vector for the subsequent fixed codebook (FCB) search.

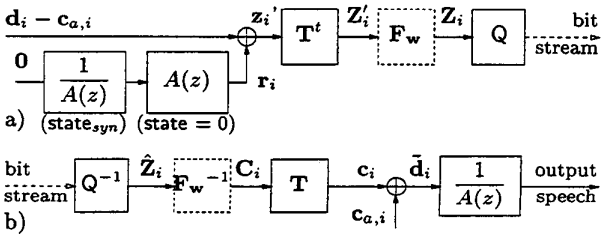

Fig. 1: Proposed TCX scheme: a) encoder, b) decoder.

Here, all filtering for the target computation is approximated in the frequency domain by a diagonal matrix $\mathbf{F_w}$ containing the sampled magnitude response of the weighted synthesis filter. Prior to the unitary frequency domain block transform of length N (given by the matrix $\mathbf{T}^t$), the ringing of the synthesis filter, transferred into the residual domain by inverse LP filtering with zero states, is added to $\mathbf{d}_i - \mathbf{c}_{a,i}$. In the frequency domain, the quantization of the target vector $\mathbf{Z}_i$ follows an AbS criterion for the innovation vector $\mathbf{C}_i$:

$$P_e = \| \mathbf{F_w}(\mathbf{Z}_i' - \mathbf{C}_i) \|^2 = \sum_{\mu=0}^{N-1} | Z_i(\mu) - F_w(\mu) C_i(\mu) |^2 \tag{1}$$

Due to the diagonal structure of $\mathbf{F_w}$, this minimization can be replaced by a direct quantization of $Z_i'(\mu) = Z_i(\mu)/F_w(\mu)$ by $C_i(\mu)$, provided that an adaptive bit allocation, con-

trolled by the speech segment's weighted psd estimated by $F_w^2(\mu)$ for $\mu = 0 \ldots N-1$, is applied.
It is important to note that, apart from approximating the filter operations in the frequency domain, the structure can still be seen as an AbS scheme: depending on the choice of weighting, the adaptive bit allocation allows for either a white or spectrally shaped reconstruction error, while the ringing of the synthesis filter is still included into the optimization. As the main difference to the TCX structure proposed in [4], the block transform is directly applied in the residual domain here. Since there is no filtering while computing the target, the final codevector $\mathbf{C}_i$ ist directly transformed into the time domain innovation $\mathbf{c}_i$, appropriate for synthesis filtering and updating the ACB. Therefore, inverse weighting operations as in [4] are not necessary. Consequently, with regard to eq. (1) and Fig. 1, the effect of the weighting by $\mathbf{F_w}$ and the inverse weighting in the decoder, which is marked in dotted boxes in Fig. 1, only refers to a proper frequency dependent gain scaling with respect to the quantizer realization and may be dropped, as the shape of the error spectrum is controlled by the adaptive bit allocation.

Simple frequency domain codebooks

We chose a scalar approach for quantizing the transform coefficients. However, to better cope with the dynamics of the input signal, a shape-gain decomposition is used:

$$P_e = g^2 \sum_{\mu=0}^{N-1} F_w^2(\mu) \left| \frac{Z_i(\mu)}{g F_w(\mu)} - \overline{C}_i(\mu) \right|^2 \rightarrow \min . \quad (2)$$

Due to the common gain g, the coefficients cannot be considered independently. To circumvent this problem, the (scalar) quantization of the shape components $\overline{C}_i(\mu)$ is repeated for all or for a preselected set of the (e.g.) 16 quantized gain values g, finally selecting the best combination.
To increase the efficiency at low transmission rates, only fractions of bits are assigned to the coefficients, namely a discrete number of quantizer levels $u(\mu)$. A corresponding allocation algorithm can be deduced from Noll's general procedure [10], regarding constraints as a maximum number of levels or an integer overall bit rate for all coefficients. From the level allocation $u(\mu)$ and the single index values $I_\mu \in \{0, \ldots, u(\mu)-1\}$ resulting from the quantization process, an overall channel index is encoded. For the coefficients with $u(\mu)$=1, a noise filling procedure is applied to suppress the effect of musical tones.
During several experiments we used nonuniform scalar quantizers with at most 20 levels and either a DCT or DFT of length N=64. We compared the results of our TCX proposal to a structure similar to [4] and found that both structures perform very similarly. During informal listening, our structure was sometimes evaluated less noisy. The DCT was preferred. In comparison to a time-domain algebraic CELP (ACELP) codebook of the same rate (36 bits for 64 shape samples), the frequency domain version was slightly preferred. Therefore, we consider the use of frequency domain codebooks within a CELP codec as an alternative of the usual time domain algebraic codebooks. Apart from the low complexity due to the direct quantization of the transform coefficients, more flexibility can be introduced into the codec design: the explicit allocation of transmission rate to the coefficients allows for more sophisticated quantization and psychoacoustic masking techniques known from audio coding. Finally, time and frequency domain codebooks may be combined, as shown in section 4.

3. COMBINED, SIGNAL DEPENDANT FA/BA LINEAR PREDICTION

In a conventional CELP codec, the model order N_p of the forward adaptive LP synthesis filter (FA-LP) is normally chosen with respect to speech signals. For other signals such as music, the spectral resolution with N_p=16 ... 18 often is not sufficient. To increase N_p, backward adaptive LP analysis (BA-LP) is necessary. In [7] and in earlier versions of [3], a CELP approach with switched FA/BA-LP is used. There, the BA mode is active mainly for music, based on a sophisticated classifier. In this mode no FA-LP parameters are transmitted. Thus the innovation coding (FCB) can profit from a higher bit rate. An alternative solution, without any change in the excitation bit rate, is to *cascade* FA- and BA-LP filters. As shown in Fig. 2, the BA-LP filter is situated in the residual domain of the FA-LP filter. In our study, we use the same adaptation technique based on the reconstructed signal as in [8, 9], using two stage computation and recursive windowing. During the experiments we found, however, that the performance for music improved by using this permanent cascade, especially at high overall transmission rates, whereas the quality of speech signals was degraded significantly.
Two reasons may affect such a high resolution LP modelling for the current frame: at first, for instationary signals, the analysis is carried out using the past output signal samples and thus may not reflect the *current* spectral characteristics. Second, the analysis is corrupted by quantization noise. In order to separate these influences, an experiment was performed where the BA analysis applied to the past *input* signal. The performance for music increased again significantly, but the degradation for speech remained.
Therefore, we concluded that the operation of the BA-LP filter should be made *adaptive*, i.e. it should be active only when yielding an improvement. For the classification, we found a straightforward but efficient measure: the prediction gains of both the FA-LP filter ($G_{p,f}$) and the FA/BA-LP cascade ($G_{p,fb}$) are computed, and the BA-LP filter is used only if $G_{p,fb} > G_{p,f}$. It should be noted that this apparently simple classifier is not an open-loop criterion. Although $G_{p,fb}$ is obtained by filtering the input signal, it is sensitive to corruptions by excessive quantization noise as well as by signal instationarities because the filter parameters are computed from the past output signal. Furthermore, the classifier does not depend on any absolute threshold selections.
Since the switching of the BA-LP filter occurs in the residual domain, it does not produce any significant noise, if the filter states are handled properly. The activity of the BA-LP filter is indicated by one bit per frame. A codec based on this LP method is very robust with regard to switched FA/BA-LP schemes such as [7], because the overall spectral shape is driven by the FA-LP filter and the BA-LP filter is not always active.

4. EXPERIMENTAL CODING SCHEME

The two techniques described above, i.e. frequency domain excitation and FA/BA-LP filtering, have been examined within a wideband CELP coding structure according to

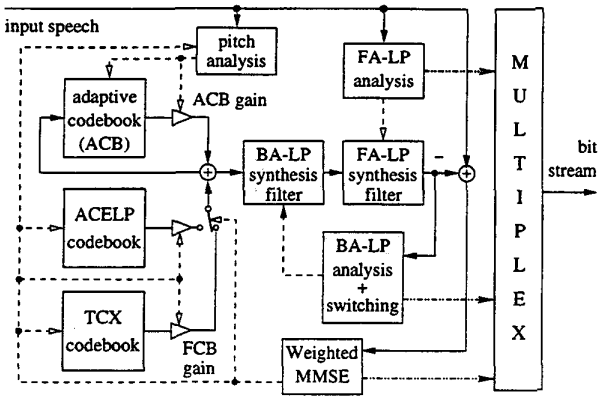

Fig. 2: Proposed wideband CELP encoder.

Fig. 2. The FA-LP parameters (model order 16) are quantized in the line spectral frequency (LSF) domain using predictive multistage split VQ with 43 bits per frame (20 ms). In each of the 5 subframes, the ACB parameters are found by a two-stage open-/closed loop pitch analysis, where the lags are encoded either absolutely by 8 or differentially by 6 bits. All ACB and FCB gains are quantized with 4 or 5 bits each; recursive prediction is used within the FCB gain quantizer. In case of a time domain excitation, an ACELP codebook with interleaved pulses and an efficient depth-first tree search is selected that uses 36 or 54 bits for a block of 64 samples. Alternatively, a DCT based frequency domain codebook of the same overall bit rate can be selected for each block. The FCB mode has to be indicated by one additional bit per subframe. The resulting bit allocation for the operation at 15.5 or 20 kbit/s is given in table 1.

5. RESULTS AND CONCLUSION

We varied the basic coding scheme to evaluate the new modules and compared the performance by informal listening. Enforcing either the ACELP or TCX mode, the quality was rated comparable. Allowing the encoder to select between the ACELP and the TCX codebook in terms of the maximum SNR, the TCX codebook is chosen in more than 50 % of the blocks. The quality gain when using such a switched excitation is significant, which is not only caused by the increase of the excitation bit rate by one bit per subframe. This was also expressed in terms of a significantly increased global SNR. Therefore and with respect to the yet simple scalar quantizer realization, we consider this FCB type as very promising. These tendencies are similar for speech and music signals. Further improvements for music can be obtained by a signal dependent adjustment of the subframe (transform) length.

	15.5 kbit/s	**20.0 kbit/s**
LPC/LSF	43	43
BA-LPC mode	1	1
ACB lag	8+6+8+6+8	8+6+8+6+8
ACB gain	5×4	5×4
FCB mode	5	5
FCB gain	5×5	5×5
FCB shape	5×36	5×54
Total # bits/frame	310	400

Table 1: Bit allocation at 15.5 and 20.0 kbit/s

For the BA-LP filter, we used a model order of 44. It is activated in about 10 ... 15 % of the frames for speech and in 50 ... 80 % for music, depending on the overall bit rate. For speech, it seems to support the FA-LP filter in voiced segments and allows to limit the FA-LP model order. Especially at the higher bit rate, the BA-LP filter gives more transparency for music signals. It should be emphasized that since all LP filter and excitation switching occurs in the residual domain, it can be performed without introducing additional degradation due to the switching itself.

Therefore, we conclude that the described techniques are appropriate to improve the performance of wideband CELP codecs. Obviously, the combination of time and frequency domain excitation and an additional BA-LP filter produce synergy effects towards a better coding performance.

6. REFERENCES

[1] J. Paulus and J. Schnitzler. "16 kbit/s Wideband Speech Coding Based on Unequal Subbands". In *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pages 651–654, 1996.

[2] J. Schnitzler. "A 13.0 kbit/s Wideband Speech Codec Based on SB-ACELP". In *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pages 157–160, 1998.

[3] P. Combescure, J. Schnitzler, K. Fischer, R. Kirchherr, C. Lamblin, A. Le Guyader, D. Massaloux, C. Quinquis, J. Stegmann, and P. Vary. "A 16, 24, 32 kbit/s Wideband Speech Codec Based on ATCELP". In *Proc. Int. Conf. Acoust., Speech, Signal Processing*, 1999.

[4] R. Lefebvre, R. Salami, C. Laflamme, and J.-P. Adoul. "High Quality of Wideband Audio Signals Using Transform Coded Excitation (TCX)". In *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pages 193–196, 1994.

[5] J.-H. Chen and D. Wang. "Transform Predictive Coding of Wideband Speech Signals". In *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pages 275–278, 1996.

[6] C.G. Gerlach. "CELP Speech Coding with Almost No Codebook Search". In *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pages (II) 109–112, 1994.

[7] S. Proust, C. Lamblin, and D. Massaloux. "Dual Rate Low Delay CELP Coding (8 kbit/s 16 kbit/s) using a Mixed Backward/Forward Adaptive LPC Prediction". In *Proc. IEEE Workshop on Speech Coding for Telecommunications*, pages 37–38, September 1995.

[8] M. Serizawa, A. Murashima, and K. Ozawa. "A 16 kbit/s Wideband CELP Coder with a High-Order Backward Predictor and its Fast Coefficient Calculation". In *Proc. IEEE Workshop on Speech Coding for Telecommunications*, pages 107–108, September 1997.

[9] A. Ubale and A. Gersho. "A Low-Delay Wideband Speech Coder at 24 Kbps". In *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pages 165–168, 1998.

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