AN ADAPTIVE MULTI RATE WIDEBAND SPEECH CODEC WITH ADAPTIVE GAIN RE-QUANTIZATION

*Christoph Erdmann**, *Peter Vary**, *Kyrill Fischer*†, *Joachim Stegmann*†,
Catherine Quinquis‡, *Dominique Massaloux*‡ *and Balázs Kövesi*‡

*Aachen University of Technology (RWTH), IND, D-52056 Aachen, Germany
†T-Nova Deutsche Telekom, D-64276 Darmstadt, Germany
‡France Télécom R&D, DIH/DIPS, F-22307 Lannion Cedex, France

ABSTRACT

This paper describes an adaptive multi rate wideband (AMR-WB) speech codec proposed for the GSM system and also for the evolving *Third Generation* (3G) mobile speech services. The coder is a multi rate SB-CELP (*Subband-Code-Excited Linear Prediction*) with five modes operating at bit rates from 24 kbit/s down to 9.1 kbit/s. Our basic approach consists of an unequal bandsplitting of the input signal into two subbands (SB). A variable rate, multi-mode ACELP coder is applied to the lower subband (0-6 kHz). The various bit rates are integrated in a common structure where the scalability is realized by exchanging the fixed excitation codebooks while leaving all other codec parameters invariant. For the GSM related modes (9.1-17.8 kbit/s), the upper subband (6-7 kHz) is coded using a very low bit rate representation based on bandwidth expansion techniques. In case of the 3G application (24 kbit/s) the upper band is coded using a 4 kbit/s ADPCM coding scheme. In addition the analysis by synthesis (AbS) coder of the lower band employs a novel *closed loop* gain re-quantization technique controlled by the character of the speech signal. Thereby the codec achieves an enhanced performance for background noise while maintaining its clean speech quality.

1. INTRODUCTION

Following the standardization of the GSM AMR *narrowband* (AMR-NB) system in 1998 [1], ETSI/3GPP conducted a feasibility study for introducing an additional wideband speech (7 kHz) mode into the AMR system [2]. The AMR-WB codec is foreseen not only for the existing GSM network, but also for future mobile radio systems of the Third Generation, where high data rate channels well above 22.8 kbit/s will be realized by packetized networks or multi timeslot configurations such as EDGE, GPRS or UMTS/IMT 2000.

The goal was to satisfy a growing market interest in a wideband speech service by providing a system that preserves or even exceeds at least the quality of the ITU-T G. 722 @ 48 kbit/s codec under the conditions of a mobile radio channel. Hence the AMR-WB system has to provide several codec modes with different balances between the source and channel coding contributions of the gross bit rate. Analogue to the AMR-NB system it features the typical AMR behaviour under dynamic channel conditions, known as a trade-off between speech quality and error robustness along with the channel quality.

In June 1999 the technical subgroups ETSI/SMG 11 and 3GPP TSG SA4 decided to start a competitive selection process. The process includes subjective qualification in April/May and selection tests in July 2000.

Finally five application scenarios A-E were identified to be relevant for the standardization. Application A and B are both meant for single timeslot operation on the GSM full rate traffic channel within its limiting gross bit rate of 22.8 kbit/s. Application A is further required to remain below 14.4 kbit/s for the source encoding to allow 16 kbit/s submultiplexing on the A_{ter} interface. The applications C, D and E refer to the so-called *3G-channels* without any additional constraints besides a maximum allowed bit rate of 32 kbit/s for application E. The speech quality for the applications A, B and C,D,E is required to be equivalent to the ITU-T G. 722 codec at 48, 56 and 64 kbit/s respectively, at most operating conditions [3].

In this article, we describe a codec proposal for the ETSI/3GPP AMR-WB standardization, which was submitted for qualification in March 2000. Our proposal uses an SB-CELP algorithm, which combines a variable rate ACELP codec in the lower band (0-6 kHz) with either bandwidth expansion or ADPCM coding of the upper band (6-7 kHz) to meet the requirements for each application. The algorithm in this paper is similar to that presented in [4]. The major difference of the lower band ACELP algorithm is that besides the rate scalability a novel signal adaptive re-quantization technique is applied for coding the fixed codebook gains. This gives a significant quality improvement in case of stationary background noise.

2. AN AMR WIDEBAND SOURCE CODEC

Our basic SB-CELP approach is based on our previous work [4, 5, 6]. Five different codec modes at overall bit rates of 9.1, 12.4, 14.2, 17.8 and 24 kbit/s are realized. The first four modes are meant for operation on the GSM full rate traffic channel, while the 24 kbit/s mode covers future applications on 3G-channels. The input signal is split into two subbands, each critically decimated, in order to allocate the available bit rate according to both the spectral distribution and the subjective importance of the subband components. We found an unequal band splitting at a cutoff frequency of 6 kHz to be a suitable solution [7]. This conclusion was motivated by the analysis of the instantaneous bandwidth of speech signals and by the spectral resolution of the auditory perception: the 6-7 kHz band corresponds to about one critical band only. A block diagram of the encoder is shown in Figure 1.

2.1. Upper band (6-7 kHz) processing

In our configuration, those spectral portions of the upper subband (6-7 kHz) which are sufficient to convey a correct subjective impression of wideband speech can be represented by coding them at a very low bit rate based on bandwidth expansion techniques. This requires only 6 bits per 20 ms frame (UB-switch closed) [4, 5]. If the overall bit budget is sufficently big (24 kbit/s mode - UB-switch open), an ADPCM coding scheme at 2 bit per sample with backward-adaptive prediction and backward-adaptive quantization (APB-AQB, [8]) is used instead.

7803-6416-3/00/$10.00©2000 IEEE

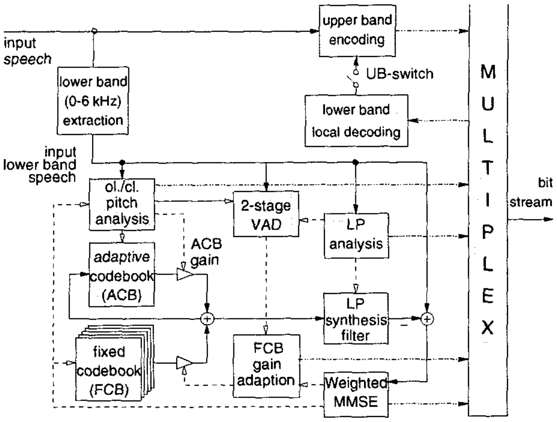

Fig. 1: Variable rate SB-CELP encoder

2.2. Lower Band (0-6 kHz) Processing

In the encoder, the input signal is bandlimited to 6 kHz and critically decimated for the lower band processing, In case of the low bit rate approach for coding the upper band the band splitting allows the lower subband (0-6 kHz) to be quantized more precisely: for example at an overall target bit rate of 12.4 kbit/s, the effective bit rate increases from $\overline{R} = 0.775$ bit per sample at a sampling rate of $f_s = 16$ kHz to $\overline{R} \approx 1$ bit per sample at $f_s = 12$ kHz.

This suggests the use of state-of-the-art ACELP (Algebraic Code-Excited Linear Prediction) techniques for coding the lower subband.

Short-Term Prediction

The short term (LP) synthesis filter coefficients are updated every 20 ms frame (240 samples at f_s=12 kHz). Look-ahead in time by 5 ms is used within the autocorrelation analysis. The quantization of the 14 LP parameters is performed in the LSF (Line Spectral Frequencies) domain and based on switched 1st order MA predictive split vector quantization, known as the *Safety-Net* approach [9], using 36 bits.

Voicing Analysis and ACB Excitation

Every 10 ms, an open-loop pitch estimate is calculated. Using this estimate, a voicing decision is taken and coded by 1 bit. Provided the 10 ms subframe is declared voiced, a constrained closed-loop adaptive codebook (ACB) search with fractional delays is performed every 5 ms [10, 7]. This procedure requires 8+6=14 bit per 10 ms for coding the pitch lags. Every 5 ms ACB-subframe, the pitch gain is nonuniformly quantized with 4 bit.

Innovative Excitation

The required rate scalability is realized by exchanging the algebraic fixed excitation codebooks and their corresponding gain codebooks while leaving the coding scheme for all the other codec parameters invariant. Thus our coding scheme exhibits a very robust behaviour against AMR-mode misdetection since a misinterpretation of the fixed excitation usually only leads to minor distortions in the reconstructed speech. Furthermore, seamless mode switching can be realized by simply changing the excitation codebooks.

Depending on the voicing mode and thus the available rate for the innovative excitation different algebraic fixed codebooks [10] are searched for the optimum innovation shape vector. A corresponding gain factor is coded using predictive scalar quantization.

Adaptive FCB Gain Re-Quantization

A novel feature of our codec is the method used to determine the FCB gain factor. This method helps to overcome unwanted gain fluctuations as they appear in CELP coders when coding pure background noise segments or speech in background noise. To determine the innovation shape $\mathbf{c}_f$ and gain g_f, AbS coding schemes usually employ a Minimum Mean Square Error (MMSE) criterion in a weighted speech domain, given by the expression

$$D = \|\mathbf{x} - g_f \, \mathbf{H} \, \mathbf{c}_f\|^2. \tag{1}$$

$\mathbf{c}_f$ and g_f are used to match the target vector $\mathbf{x}$, when filtered with the impulse response of the weighted synthesis filter $\mathbf{H}$. The minimization of (1) with respect to g_f leads to the optimal uncoded gain value for a given codebook vector $\mathbf{c}_f$ given by:

$$g_f = \frac{\mathbf{x}^t \mathbf{H} \mathbf{c}_f}{\mathbf{c}_f^t \mathbf{H}^t \mathbf{H} \mathbf{c}_f} \quad . \tag{2}$$

Following (2) the coded gain strongly depends on the matching between the waveform of the coded innovation $g_f \, \mathbf{H} \, \mathbf{c}_f$ and the target vector $\mathbf{x}$. Thus, in noise like segments of the signal, where most of the time the match is poor, these varying waveform matching abilities cause the gain value to fluctuate, what becomes audible as an annoying artifact, known as "swirling".

In a recent example [11] of a multi-mode coder, for the unvoiced mode as well as for background noise frames encoded with the unvoiced mode, the gain was found to be more important than the exact waveform match by setting g_f so that the energy level of the LP residual $\mathbf{r}$ is matched instead of using (2). Another example of a single-mode coder is given in [12], where the adaptive codebook is used regardless of the current speech frame's character. An adaptive error criterion is proposed which provides seamless cross-fading between waveform and energy matching. This criterion is given by:

$$D = (1-\alpha)\cdot\|\mathbf{c}_f\|^2\cdot(g_f - \hat{g})^2 + \alpha\cdot(\|\mathbf{r}\| - \hat{g}\cdot\|\mathbf{c}_f\|)^2. \tag{3}$$

Even for a multi-mode coder like ours, we found it more promising to apply such an adaptive error criterion, instead of relying on either waveform or energy matching according to the voiced/unvoiced mode decision. Due to the mode dependent usage of the adaptive codebook, any contribution $g_a \cdot \mathbf{c}_a$ of the adaptive codebook to the excitation has to be considered in addition. Therefore, we formulate a modified error criterion, given by:

$$\begin{aligned} D_{mod} = (1-\alpha)\cdot\|\mathbf{c}_f\|^2\cdot(g_f - \hat{g})^2 \\ + \alpha\cdot(\|\mathbf{r}\| - \|\hat{g}\cdot\mathbf{c}_f + g_a\cdot\mathbf{c}_a\|)^2. \end{aligned} \tag{4}$$

Note that g_f denotes the quantized closed loop gain factor with respect to the pure waveform matching criterion from (2). Hence the gain codebook is searched once again for the entry $\hat{g}$ that minimizes (3). The deviation of the re-quantized gain factor $\hat{g}$ from g_f depends on the value of α. Thus, the adaption of the balance factor α is crucial for this approach.

We found it beneficial to base the adaption on two characteristic features of the input speech as they are:

- the voicing criterion v_{crit} which is also responsible for the voiced/unvoiced mode decision
- a measure indicating the presence of a stationary noise like segment

In case of speech activity or instationary noise segments we choose the adaption factor α as a function of the averaged voicing criterion $\bar{v}_{crit}$ as proposed in [12]. The voicing criterion itself is based on the open loop pitch estimate τ_{ol}, given by:

$$v_{crit} = \frac{\sum_{i=0}^{L-1} s(i)\, s(i-\tau_{ol})}{\sqrt{\sum_{i=0}^{L-1} s^2(i) \cdot \sum_{i=0}^{L-1} s^2(i-\tau_{ol})}}. \qquad (5)$$

Since even $\bar{v}_{crit}$ is closely related to the periodicity of the current speech frame $\mathbf{s}$, it tends to fluctuate for segments of pure stationary background noise. Using the mixed criterion from (4) leads to an indifferent gain adaption where in fact pure energy matching is required. Therefore, in case of stationary noise only segments we choose to rely exclusively on the energy matching criterion by setting $\alpha = 1$. For detecting those regions of stationary background noise we use a simple yet effective voice activity detection (VAD) which monitors variations in the energy level and the LPC spectrum with respect to the voicing criterion. When the VAD indicates that the frame does not contain speech, the presence of stationary background noise is assumed.

Perceptual Weighting and Postfiltering

In the adaptive and fixed codebook search processes, an adaptive perceptual weighting filter is used. Adaptive postfiltering is applied to the synthesized lower band speech.

Decoder

At the decoder, the synthesis filter bank interpolates and superposes the decoded lower and upper band signals, yielding the wideband output signal. The delay of the analysis/synthesis filter bank amounts to 10 ms.

3. RESULTS

We conducted informal listening tests in our laboratory to evaluate especially the background noise performance of the 12.4 kbit/s coder in comparison with our former approach at 13 kbit/s using conventional FCB gain coding [4]. Clean speech and speech with 15 dB car and street noise were tested in a DCR manner, involving experienced listeners. The quality for clean speech was rated equal with a slight preference for the new coder. For both car and street noise a significant preference for the new coder was observed. Thus, we concluded that the proposed adaptive FCB gain re-quantization clearly improves performance for background noise while maintaining the clean speech quality.

Further we added state of the art channel coding schemes for the 9.1, 12.4, 14.2 and 17.8 kbit/s mode based on convolutional codes and unequal error protection aiming a gross data rate of 22.8 kbit/s, corresponding to the GSM full rate channel. Using an appropriate AMR mode adaption the codec remains stable even for a channel quality below $C/I = 10\,\text{dB}$.

The algorithm is presently undergoing formal subjective listening testing within the ETSI/3GPP AMR-WB qualification. The results will be presented during the presentation.

4. CONCLUSIONS

A variable rate wideband codec based on SB-CELP has been proposed for coding wideband speech at various bit rates from 24 kbit/s down to 9.1 kbit/s. Using state of the art channel coding for each mode the codec features the AMR principal and remains stable at low to medium error conditions. Therefore it is appropriate for operation on the GSM full rate channel as well as for high data rate 3G-channels. In this paper we have presented a novel adaptive gain re-quantization technique. Informal listening tests have shown that the new approach helps to overcome the CELP typical problems on coding background noise.

The audio quality of our codec will be demonstrated at the conference.

5. ACKNOWLEDGEMENTS

The authors wish to thank Joachim Hagenauer, Markus Kaindl and Thomas Hindelang from the LNT / TU Munich for fruitfull discussions and their support in implementing the channel coding scheme.

6. REFERENCES

[1] E. Ekudden, R. Hagen, I. Johansson, and J. Svedberg, "The Adaptive Multi-Rate Speech Coder", in *Proc. IEEE Workshop on Speech Coding*, Porvoo, Finland, June 1999, pp. 117–119.

[2] ETSI SMG 11, "Adaptive Multi-Rate Wideband (AMR-WB) Feasibility Study Report", Version 1.0.00.2.0, Tdoc SMG 265/99, June 1999.

[3] ETSI SMG 11 and 3GPP TSG-S4, "AMR Wideband Performance Requirements", Version 2.0, Tdoc S4/SMG 11 (00)00173, Feb. 2000.

[4] J. Schnitzler, C. Erdmann, P. Vary, K. Fischer, J. Stegmann, C. Quinquis, D. Massaloux, and C. Lamblin, "Wideband Speech Coding for the GSM Adaptive Multi Rate System", in *Proc. 3rd ITG Conference, Source and Channel Coding*, Munich, Germany, Jan. 2000, pp. 325–329.

[5] J. Schnitzler, "A 13.0 kbit/s Wideband Speech Codec Based on SB-ACELP", in *Proc. ICASSP*, Seattle, WA, USA, 1998, IEEE, pp. 157–160.

[6] 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. ICASSP*, Phoenix, AZ, USA, 1999, IEEE, pp. (I) 5–8.

[7] J. Paulus and J. Schnitzler, "Wideband Speech Coding for the GSM Fullrate Channel ?", in *Proceedings ITG-Fachtagung Sprachkommunikation*, Frankfurt am Main, 1996, pp. 11–14.

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

[9] T. Eriksson, J. Linden, and J. Skoglund, "Exploiting Interframe Correlation in Spectral Quantization – A Study of Different Memory VQ Schemes", in *Proc. ICASSP*, Atlanta, GA, USA, 1996, IEEE, pp. 765–768.

[10] R. Salami, C. Laflamme, J.P. Adoul, A. Kataoka, S. Hayashi, T. Moriya, C. Lamblin, D. Massaloux, S. Proust, P. Kroon, and Y. Shoham, "Design and description of CS-ACELP: A Toll Quality 8 kb/s Speech Coder", *IEEE Trans. Speech and Audio Processing*, vol. 6, no. 2, pp. 116–130, Mar. 1998.

[11] E. Paksoy, A. McCree, and V. Viswanathan, "A Variable-Rate Multimodal Speech Coder with Gain-Matched Analysis-by-Synthesis", in *Proc. ICASSP*, Munich, Germany, 1997, IEEE, pp. 755–758.

[12] R. Hagen and E. Ekudden, "An 8 kbit/s ACELP Coder with Improved Background Noise Performance", in *Proc. ICASSP*, Phoenix, AZ, USA, 1999, IEEE, pp. 25–28.