A CANDIDATE PROPOSAL FOR A 3GPP ADAPTIVE MULTI-RATE WIDEBAND SPEECH CODEC

Christoph Erdmann[†], *Peter Vary*[†], *Kyrill Fischer*[‡],
Wen Xu[◇], *Matthias Marke*[◇], *Tim Fingscheidt*[◇], *Imre Varga*[◇], *Markus Kaindl*[▷],
Catherine Quinquis[⋆], *Balázs Kövesi*[⋆] *and Dominique Massaloux*[⋆]

[†]Aachen University of Technology (RWTH), IND, D-52056 Aachen, Germany

[‡]T-Nova Deutsche Telekom, D-64276 Darmstadt, Germany

[◇]Siemens AG, Mobile Phones, D-81675 Munich, Germany

[▷]Munich University of Technology (TUM), LNT, D-80290 Munich, 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 speech codec is based on SB-CELP (*Splitband-Code-Excited Linear Prediction*) with five modes operating bit rates from 24 kbit/s down to 9.1 kbit/s. The respective channel coding schemes are based on RSC (Recursive Systematic Code) and UEP (Unequal Error Protection). Both, source and channel codec are designed as homogenous as possible to guarantee robust transmission on current and future mobile radio channels.

1. INTRODUCTION

Following the standardization of the GSM AMR *narrowband* (AMR-NB) system in 1998 [1, 2], ETSI/3GPP conducted a feasibility study for introducing an additional wideband speech (7 kHz) mode into the AMR system [3]. The AMR-WB codec is intended 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 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. Analoguous to the AMR-NB system it features the typical AMR behavior 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 included subjective qualification in April/May 2000 and selection tests in September 2000.

Finally, four application scenarios A,B,C, and E were identified to be relevant for the standardization. Applications 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. Applications C 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/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 [4].

In this article, we describe a codec proposal for the 3GPP AMR-WB standardization, which was submitted for qualification in March 2000 and became a candiate for selection. 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.

2. AMR WIDEBAND SPEECH CODEC

Our SB-CELP approach is based on our previous work [5, 6, 7, 8]. 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 [9]. This conclusion was motivated by the analysis of the instantaneous bandwidth of speech signals and by the spectral resolution of 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) [6, 7]. If the overall bit budget is sufficiently 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, [10]) is used instead.

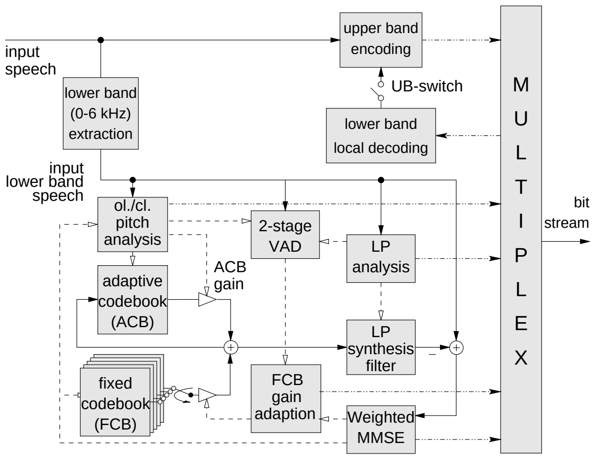

Fig. 1: Variable rate SB-CELP encoder

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 [11], 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. The underlying voicing criterion is given by

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

Provided the 10 ms subframe is declared voiced, a constrained closed-loop adaptive codebook (ACB) search with fractional delays is performed every 5 ms [12, 9]. 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 behavior 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 exchanging the excitation codebooks.
Depending on the voicing mode and thus the available rate for the innovative excitation, different algebraic fixed codebooks [12] are searched for the optimum innovation shape vector every 5, 2.5, 1.6 or 0.83 ms. A corresponding gain factor is coded using predictive scalar quantization. To enhance the performance when coding background noise segments or speech in background noise, adaptive gain requantization is applied [13].

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. It consists of a long term post-filter through re-analyzing the pitch of the decoded speech, a short term post-filter, tilt compensation and automatic gain control [14].

2.3. 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 2.25 ms.

2.4. Bit-Allocation

Parameter	**Subframe 1**		**2**		**3**		**4**	
	v	uv	v	uv	v	uv	v	uv
all modes								
LSF	36							
V/UV	1		0		1		0	
ACB	8	0	6	0	8	0	6	0
g_{acb}	4	0	4	0	4	0	4	0
9.1 kbit/s								
FCB	20		20		20		20	
g_{fcb}	4		4		4		4	
UBR	2		1		2		1	
$\sum$	182							
12.4 kbit/s								
FCB	36	3·13	36	3·13	36	3·13	36	3·13
g_{fcb}	4	3·4	4	3·4	4	3·4	4	3·4
UBR	2		1		2		1	
$\sum$	248							
14.2 kbit/s								
FCB	45	3·16	45	3·16	45	3·16	45	3·16
g_{fcb}	4	3·4	4	3·4	4	3·4	4	3·4
UBR	2		1		2		1	
$\sum$	284							
17.8 kbit/s								
FCB	2·28	6·9	2·28	6·9	2·28	6·9	2·28	6·9
g_{fcb}	2·5	6·4	2·5	6·4	2·5	6·4	2·5	6·4
UBR	2		1		2		1	
CRC	2 ... 6							
$\sum$	356							
24.0 kbit/s								
FCB	2·34	3·25	2·34	3·25	2·34	3·25	2·34	3·25
g_{fcb}	2·5	3·15	2·5	3·5	2·5	3·5	2·5	3·5
ADPCM	80							
CRC	2 ... 4							
$\sum$	480							

Table 1: Overall Bit-Allocation (v: voiced; uv: unvoiced)

3. AMR WIDEBAND CHANNEL CODEC

After speech encoding, the generated bits are protected using recursive systematic convolutional (RSC) codes [15]. As required by the standardization group, convolutional codes with constraint length 5 for applications A/B and 7 for application C are employed. The generator polynomials are defined in [16]. Since the bits are not of equal perceptual importance, different bit errors after transmission have differently severe impacts. Therfore, a scheme of unequal error protection (UEP) has been designed by appropriately puncturing the codebits. The most important bits are highly protected whereas less important bits are less protected. Like in the existing GSM AMR-NB codec, the speech encoded bits are classified into 3 classes: Ia, Ib und II. Prior to the convolutional coding, the most important class Ia bits are protected by a 6 bit CRC generated by the cyclic generator polynomial:

$$g(D) = D_6 + D_5 + D_3 + D_2 + D_1 + 1. \qquad (2)$$

Evaluation of this CRC at the receiver side is used to set a bad frame indicator (BFI) which controls the speech decoder error concealment.
The GSM full rate (FR) traffic channel interleaver is used to interleave the channel coded bits into 8 time slots. The speech bits are channel-coded such that the gross bit rate of 22.8 kbit/s is generated. The main decoding routine is a Viterbi algorithm. For the purpose of the AMR codec mode adaption on the GSM FR channel 8 bits are reserved, which ensures the usage of AMR narrowband blockcode for in-band signalling as defined in [16]. Mode detection and channel estimation can be implemented as described in [17].

4. PERFORMANCE

In the qualification phase subjective listening tests were performed. Following a common test plan [18], each candidate proposal was assessed individually and cross-checked by other participants. Figure 2, 3 and 4 summarize the results for clean speech, car and street noise with static channel conditions.
For application A and B the candidate was tested on GSM full rate (FR) traffic channel with constant C/I (the error conditions EC4...19 correspond to a *typical urban* channel profile with ideal frequency hopping at $C/I = 4 \ldots 19\,dB$).
For application E, instead of an explicit channel simulation, the candidate was exposed to frame erasures (FER) and residual bit errors (RBER). The different RBER patterns consider two classes of bit sensitivity with accordingly stronger or weaker error protection (e.g. 1.0%/0.1%: 1.0% RBER in the less sensitive bits, 0.1% RBER in the more sensitive bits).
Our configuration for application A included the 14.2 kbit/s mode at *NoErrors* and 13 dB C/I, 12.4 kbit/s at 10 dB C/I and 9.1 kbit/s at 7 and 4 dB C/I. For application B this configuration is extended by the 17.8 kbit/s mode at *NoErrors*, 19 and 16 dB C/I. For application E the 24.0 kbit/s mode is used in all conditions.
The results can be summarized as follows. For clean speech the performance in the FR channel is equivalent to G. 722 @ 56 kbit/s and better than G. 722 @ 48 kbit/s down to 10 dB C/I, while the relative quality decrease due to poor channel conditions is superior to the decrease demonstrated by the GSM EFR. In application E, the performance is

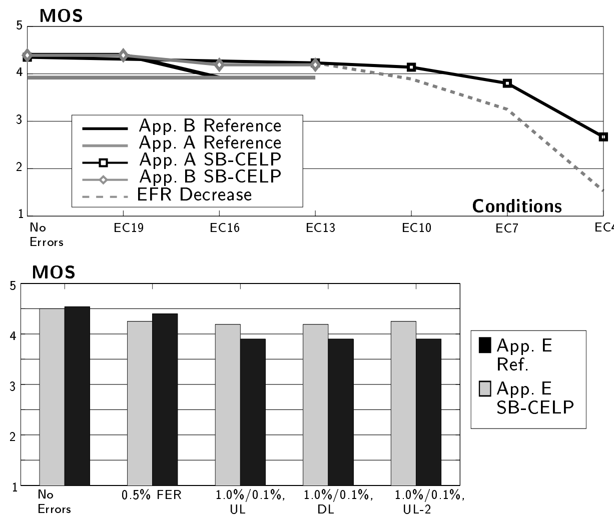

Fig. 2: Performance in clean speech

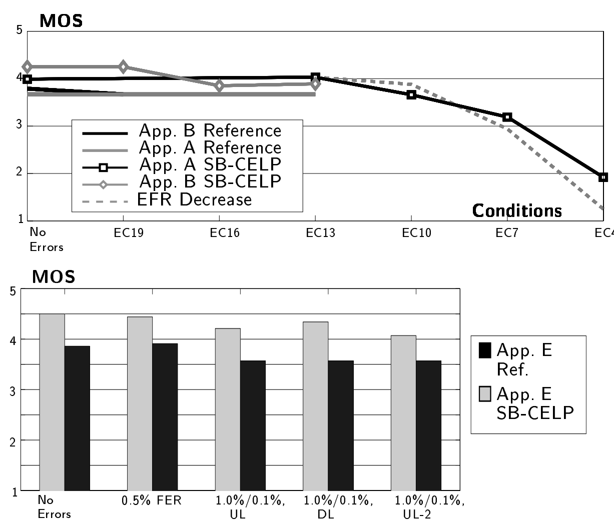

Fig. 3: Performance in car noise

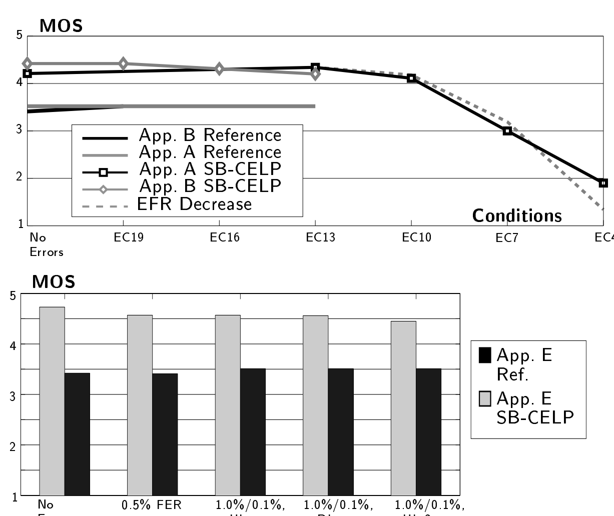

Fig. 4: Performance in street noise

equivalent to G. 722 @ 64 kbit/s and 56 kbit/s for error free conditions and 0.5% FER, respectively. For residual bit errors the quality is superior to G. 722 @ 48 kbit/s. In background noise conditions (car and street noise) the performance in the FR channel is significantly increased compared to the G. 722 @ 56 kbit/s and 48 kbit/s, while the robustness under poor channel conditions is still compareable to GSM EFR. In application E, the performance is clearly superior to the reference in all error conditions.

5. CONCLUSIONS

The AMR-WB codec was designed to meet a challenging set of requirements. The initial task to provide robust transmission of wideband speech on the GSM FR channel was expanded by the additional task to be also applicable to future 3G channels. Therefore, we proposed a variable rate wideband speech codec based on SB-CELP that covers 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 showed very satisfying performance for clean speech and background noise on GSM FR- and 3G channels. Thus, it was able to meet all requirements of the 3GPP AMR-WB qualification and became a candidate for selection. In the selection phase this candidate was assessed by international, formal listening tests [19] and proved its performance while beeing compliant with the design constraints.

6. ACKNOWLEDGEMENTS

The authors wish to thank Frederic Gabin, Aurelie Gervais and William Navarro from Nortel Networks for their support in the selection phase processing.

7. 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, *GSM 06.71: Digital cellular telecomunications system (Phase 2+); Adaptive Multi-Rate (AMR); Speech Processing Functions; General Description*, European Telecommunications Standards Institute, 1998.

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

[4] ETSI SMG 11 and 3GPP TSG-S4, "AMR Wideband Performance Requirements", Version 2.2, Tdoc S4/SMG 11 (00)00321, June 2000.

[5] J. Paulus, *Codierung breitbandiger Sprachsignale bei niedriger Datenrate*, PhD thesis, Aachener Beiträge zu digitalen Nachrichtensystemen, Publ. P. Vary, Bd. 6, 1997.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

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

[11] 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.

[12] 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.

[13] C. Erdmann, P. Vary, K. Fischer, J. Stegmann, C. Quinquis, D. Massaloux, and B. Kövesi, "An Adaptive Multi Rate Wideband Speech Codec with Adaptive Gain Re-Quantization", in *Proc. IEEE Workshop on Speech Coding*, Delavan, Wisconsin, Sept. 2000, pp. 145–147.

[14] D. Massaloux and S. Proust, "Spectral Shaping in the Proposed ITU-T 8 kb/s Speech Coding Standard", in *Proc. IEEE Workshop on Speech Coding for Telecommunications*, Annapolis, Maryland, USA, 1995, pp. 9–10.

[15] W. Xu, S. Heinen, M. Adrat, P. Vary, T. Hindelang, M. Schmautz, and J. Hagenauer, "An adaptive multirate speech codec proposed for the GSM", in *Proc. 3rd ITG Conference, Source and Channel Coding*, Munich, Germany, Jan. 2000, pp. 51–56.

[16] ETSI, *GSM 05.09: Digital cellular telecomunications system (Phase 2+); Channel Coding*, European Telecommunications Standards Institute, July 2000.

[17] ETSI, *GSM 05.09: Digital cellular telecomunications system (Phase 2+); Link Adaptation*, European Telecommunications Standards Institute, Aug. 2000.

[18] 3GPP TSG-S4 / ETSI SMG11, "Test Plans for the AMR Wideband Qualification Phase (AMR-WB-8a)", Version 1.1, Tdoc S4/SMG11 (00)00162R, Feb. 2000.

[19] 3GPP TSG-S4, "AMR Wideband Selection Test Plan (AMR-WB-8b)", Version 1.0 (Final), Tdoc S4 (00)00355, Sept. 2000.