AN EMBEDDED SCALABLE WIDEBAND CODEC BASED ON THE GSM EFR CODEC

Peter Jax, Bernd Geiser, Stefan Schandl [†], Hervé Taddei [‡], and Peter Vary

Institute of Communication Systems and Data Processing (ind)
RWTH Aachen University, Germany
{jax|geiser|vary}@ind.rwth-aachen.de

[†]Siemens AG, Vienna, Austria
stefan.schandl@siemens.com

[‡]Siemens AG, Munich, Germany
herve.taddei@siemens.com

ABSTRACT

We present a technique to extend *narrowband* (NB) speech communication systems, using e.g. the GSM *enhanced full rate* (EFR) codec [1], with *wideband* (WB, 50–7000 Hz) capability. The limited acoustic bandwidth of narrowband speech coding is extended using a fairly coarse description of the missing high frequency band (3.4–7 kHz) in terms of temporal and spectral envelopes. The high-band parameters are quantized, transmitted and then used at the receiver side to regenerate the high frequency components. The parameter encoding is done by applying split vector quantization in a transformed domain. This quantization scheme can be scaled to match any given target bit rate. Several example configurations have been implemented and tested in MUSHRA-style listening tests.

1. INTRODUCTION

When taking a closer look at todays' wideband speech or audio coding standards like the *adaptive multirate wideband* (AMR-WB) codec [2], it can be observed that certain parameters of the highest frequency subband are often extrapolated from lower frequency components. This can be interpreted as a kind of *bandwidth extension* (BWE, e.g. [3], [4]) supported by a rather low amount of side information. Examples for speech and audio coding with BWE techniques are the Extended AMR-WB+ [5] and Enhanced aacPlus [6] codecs. Several other proposals exist in literature, e.g., [7] – [10].

In the AMR-WB codec standard [2], the *extension band* (EB) components (6.4–7 kHz) are encoded and decoded using *linear prediction coding* (LPC) techniques. The encoder performs an LPC analysis of the input signal, and the linear prediction coefficients and subframe gains of the residual signal are encoded. In the decoder the EB residual signal is artificially generated, and the transmitted gain factors as well as the reconstructed LPC synthesis filter are applied.

The AMR-WB concept has been significantly extended in the AMR-WB+ codec [5]. Here, the extension band is much larger (4–8 kHz if the sampling rate is 16 kHz), and more side information (LPC coefficients and gain factors) is transmitted to support the BWE in the decoder.

In the Enhanced aacPlus standard [6] the *spectral band replication* (SBR) technique is used. The wideband speech or audio signal is split into frequency subbands by a 64-channel QMF filterbank. For the high frequency filterbank channels, parametric coding of the subband signal components is employed using several detectors and estimators to control the bitstream contents.

In this paper we consider an embedded wideband coding concept. The narrowband frequency components (50–3400 Hz) are encoded by a narrowband codec, using common techniques such as *code excited linear prediction* (CELP). The high frequency band (3.4–7 kHz) is synthesized by BWE in the receiver, supported by a certain amount of side information. Throughout the paper we will use the GSM EFR codec (12.2 kbit/s) as an example narrowband codec. However, the proposed wideband coding concept can be applied together with any similar narrowband codec as well.

The wideband "add-on" extends the acoustic bandwidth of the narrowband output signal of the EFR codec using only a coarse description of the EB components (3.4–7 kHz). This coarse description comprises time and frequency envelopes which are extracted from the EB components of the original wideband speech signal every 20 ms. Extraction and scalable quantization of these parameters will be described in Sec. 2. In the receiver the EB signal components are synthesized by shaping the time and frequency envelopes of an artificially generated excitation signal, see Sec. 3.

A major difference to previous approaches for wideband coding (like the AMR-WB or AMR-WB+) or embedded WB coding is that we do not use any LPC techniques to do the frequency envelope shaping. Instead of a conventional all-pole LPC synthesis filter we use a linear-phase *finite impulse response* (FIR) filter. Therefore, the amount of ringing artifacts, clicks, crackles etc. that may occur with strongly time-variant filters is much lower. We have taken special care to produce smooth transitions in time and frequency domain.

2. BWE TRANSMITTER

Fig. 1 depicts the encoder side of the wideband "add-on". It is assumed that the wideband input speech $s_{\mathrm{wb}}(k)$ has a sampling frequency of 16 kHz. The upper signal path includes the conventional narrowband encoding, e.g. by the EFR codec. To reduce the sampling rate before the NB encoder, low-pass filtering and decimation is applied. The lower signal path comprises the parameter extraction for both the time and the frequency envelope of the EB signal components. First, the wideband input signal $s_{\mathrm{wb}}(k)$ is band-pass filtered (3.4–7 kHz) to isolate the EB signal components $s_{\mathrm{eb}}(k)$. Every 20 ms, time and frequency envelopes of the EB signal $s_{\mathrm{eb}}(k)$ are extracted synchronously to the encoding of the narrowband signal. The envelope extraction will be described in the next two sections.

2.1. Extraction of the Time Envelope

To determine the time envelope, the EB part of the speech signal (of length 20 ms, i.e., $L = 320$ samples) is subdivided into $N_{\mathrm{T}} = 10$ overlapping subframes. This yields a sufficient time resolution which ensures a good reproduction of stop consonants and

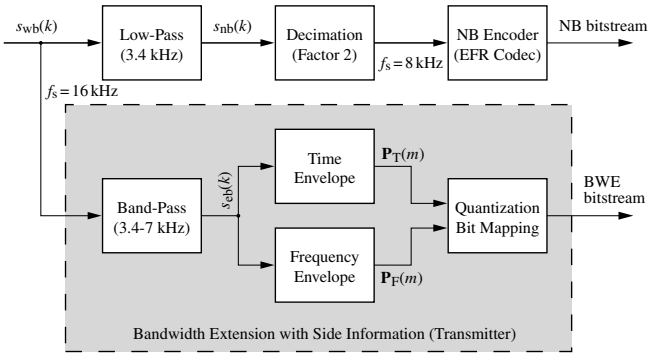

Fig. 1. BWE transmitter with parameter extraction and quantization.

plosives in speech signals. The subframes, which are indexed by $n \in \{0, \dots N_{\mathrm{T}} - 1\}$, are constructed using half overlapping Hann windows $w_{\mathrm{T}}(k)$ with a length of $L_{\mathrm{T}} = 2\,(L/N_{\mathrm{T}} - 1) + 1$. The time envelope value for the n-th subframe of the m-th signal frame is defined by the subframe energy

$$T_{m,n} = \sum_{k=0}^{L_{\mathrm{T}}-1} w_{\mathrm{T}}(k) \cdot \big(s_{\mathrm{eb}}(k + (m + n/N_{\mathrm{T}})\, L)\big)^2. \qquad (1)$$

The ten subframe energies are converted into decibels and thus form the 10-dimensional time envelope parameter vector $\mathbf{P}_{\mathrm{T}}(m) = 10\,\log_{10}\left[T_{m,0}, \dots T_{m,9}\right]^{\mathrm{T}}$.

2.2. Extraction of the Frequency Envelope

The frequency envelope of the EB signal components is computed twice per signal frame, i.e., $N_{\mathrm{F}} = 2$. To obtain a frequency domain representation, overlapping signal segments are transformed via DFT every 10 ms. The slightly asymmetric analysis window is constructed by starting with the rising slope of a 288-tap Hann window, followed by the falling slope of a 224-tap Hann window. Thus, the resulting window $w_{\mathrm{F}}(k)$ has the length $L_{\mathrm{F}} = 288/2 + 224/2 = 256$. With a lookahead of 32 samples and a lookback of 64 samples, the maximum of the window function is exactly in the center of the current subframe of length $L/N_{\mathrm{F}} = 160$.

The EB spectrum for the l-th subframe, $l \in \{0, 1\}$, of the m-th frame is determined by

$$S_{\mathrm{eb}}(m, l, \mu) = \sum_{k=0}^{L_{\mathrm{F}}-1} w_{\mathrm{F}}(k)\, s_{\mathrm{eb}}\big(k + (m + l/N_{\mathrm{F}})\, L\big) \cdot \mathrm{e}^{-\mathrm{j}\,\frac{2\pi}{L_{\mathrm{F}}} k\mu}, \qquad (2)$$

where $\mu \in \{0, \dots L_{\mathrm{F}} - 1\}$ is the frequency index.

The frequency envelope is defined by the powers of $N_{\mathrm{sb}} = 10$ subbands with equal bandwidths in the frequency range between 3.4 kHz and 7 kHz. The power $F_{m,l,\nu}$ of the ν-th subband ($\nu \in \{0, \dots 9\}$) in the l-th subframe of the m-th frame is obtained by weighted summation of the periodogram

$$F_{m,l,\nu} = \sum_{\mu=0}^{L_W-1} W_{\mathrm{F}}(\mu) \cdot \big(S_{\mathrm{eb}}(m, l, \mu + \mu_c + \nu\,(L_W - 1)/2)\big)^2. \qquad (3)$$

The frequency domain window $W_{\mathrm{F}}(\mu)$ is a Hann window with a length of $L_W = 11$. With the offset $\mu_c = 53$ the centers of the ten subbands are at the frequencies 3450 Hz + $(1/2 + \nu)$ 375 Hz.

The 20-dim. frequency envelope vector for each frame contains the subband powers of both subframes in decibels, i.e., $\mathbf{P}_{\mathrm{F}}(m) = 10\,\log_{10}\left[F_{m,0,0}, \dots F_{m,0,9}, F_{m,1,0}, \dots F_{m,1,9}\right]^{\mathrm{T}}$.

2.3. Quantization

For each signal frame of 20 ms we now have a total of 30 parameter values (10 for the time envelope and 20 for the frequency envelopes) to be jointly quantized and transmitted within the BWE bitstream. All of the parameters are gathered in a single vector $\mathbf{P}(m) = \left[\mathbf{P}_{\mathrm{T}}^{\mathrm{T}}(m), \mathbf{P}_{\mathrm{F}}^{\mathrm{T}}(m)\right]^{\mathrm{T}}$. Investigations with typical speech data showed that there is strong correlation between the parameters within $\mathbf{P}$. Therefore, a quantization scheme should be used that exploits the strong dependencies, yet with reasonable computational complexity. To achieve this trade-off we propose to use split vector quantization (VQ) in the transform domain of a (long-term) *principal component analysis* (PCA) [11]. This concept has high flexibility, good performance, and can easily be tuned for different bit rates. The individual building blocks will be described in the following.

2.3.1. Principal Component Analysis

The task of a PCA is to find a transformation matrix $\mathbf{A}$ which rotates the parameter coordinate system such that the vector elements in the rotated vector space are mutually uncorrelated and sorted by decreasing variance. The transformed vectors are labeled by $\mathbf{y}$. Mathematically, the rotation is performed by linear transformation with the orthonormal matrix $\mathbf{A}$, i.e., $\mathbf{y} = \mathbf{A}^{\mathrm{T}}\,\mathbf{P}$. The actual quantization of the parameter vector takes place in the transform domain. Thereby, at least for highly correlated parameter vectors, most of the correlation can be utilized to increase the quantizer performance. At the receiver the decoded vectors $\hat{\mathbf{y}}$ have to be transformed back into the parameter domain, $\hat{\mathbf{P}} = \mathbf{A}\,\hat{\mathbf{y}}$, to obtain the decoded envelope parameters $\hat{\mathbf{P}}_{\mathrm{T}}$ and $\hat{\mathbf{P}}_{\mathrm{F}}$.

The transformation matrix $\mathbf{A}$ is determined by evaluating the long-term statistics of the parameter vector $\mathbf{P}$ in a training phase. Parameter vectors are computed for a large training data base. These training vectors are then used to estimate the long-term covariance matrix $\hat{\mathbf{C}} = \mathrm{E}\big\{(\mathbf{P} - \mathrm{E}\{\mathbf{P}\})(\mathbf{P} - \mathrm{E}\{\mathbf{P}\})^{\mathrm{T}}\big\}$, where $\mathrm{E}\{\cdot\}$ denotes the expectation operation. The columns of the transformation matrix $\mathbf{A}$ are constructed from the eigenvectors of the estimated covariance matrix $\hat{\mathbf{C}}$. The eigenvectors (columns in $\mathbf{A}$) are sorted in decreasing order of the corresponding eigenvalues, i.e., the variances of the vector elements in $\mathbf{y}$ will be decreasing accordingly. We will see in the next section that some of the elements of $\mathbf{y}$ will not be quantized. The corresponding columns of $\mathbf{A}$ can be omitted such that, in general, the dimension of the vectors $\mathbf{y}$ will be smaller than 30.

2.3.2. Split Vector Quantization in the Transform Domain

The split *vector quantization* (VQ) described in the sequel takes place in the PCA transform domain. The principle is that the transformed vector $\mathbf{y}$ is split into sub-vectors which are then independently quantized using LBG-trained VQs [12]. That is, the transformed parameter vector is split as $\mathbf{y} = \left[\mathbf{y}_1^{\mathrm{T}}, \dots \mathbf{y}_{N_{\mathrm{VQ}}}^{\mathrm{T}}\right]^{\mathrm{T}}$, where N_{VQ} is the number of splits. The dimension of the i-th sub-vector $\mathbf{y}_i$ is labeled by $d_i = \dim \mathbf{y}_i$, and there are r_i bits per frame allocated to that sub-vector. The bit allocation has to be specified such that the sum of all r_i is equal to the available number of bits per frame.

The dimensions d_i of the sub-vectors as well as the assigned r_i bits to quantize each sub-vector have been optimized such that the expectation of the summed square quantization error is minimized. For scalar quantization, i.e., for sub-vectors with the dimension $d_i = 1 \,\forall\, i \in \{1, \dots N_{\mathrm{VQ}}\}$, this task can be fulfilled by the well-known reverse-waterfilling procedure, taking into account the overall amount of available bits per frame and the variances of the PCA transformed vectors. We have optimized the *non-scalar* splits

and bit allocations manually by a similar strategy — the resulting bit allocations are listed in Table 1 for different gross bit rates. The sub-vector dimensions d_i have been chosen such that the sizes of the VQ codebooks do not exceed 64 entries.

Table 1. Bit allocations of the split-VQs for the investigated data rates. The frame rate is 50 per second.

split-VQ no.	data rate [bit/s] on top of NB codec				
	300	600	1000	1500	2000
i	d_i r_i	d_i r_i	d_i r_i	d_i r_i	d_i r_i
1	5 6	1 4	1 5	1 6	1 6
2		1 2	2 5	2 6	1 4
3		7 6	4 6	2 5	2 6
4			9 4	4 6	2 6
5				3 3	2 4
6				9 4	4 6
7					6 6
8					5 2
sum	5 6	9 12	16 20	21 30	23 40

3. BWE RECEIVER

The BWE receiver is shown in Fig. 2. The upper signal path includes the narrowband (EFR) decoder, followed by upsampling of the signal to 16 kHz sampling frequency and low-pass filtering. The lower signal path contains the synthesis of the EB signal components. This procedure starts by generating an excitation signal $u_{\mathrm{eb}}(k)$ (see Sec. 3.1). The decoded BWE parameters $\hat{\mathbf{P}}_{\mathrm{T}}(m)$ and $\hat{\mathbf{P}}_{\mathrm{F}}(m)$ are then used to shape the time and frequency envelopes of $u_{\mathrm{eb}}(k)$ according to the properties of the original EB components (see Sec. 3.2 and 3.3).

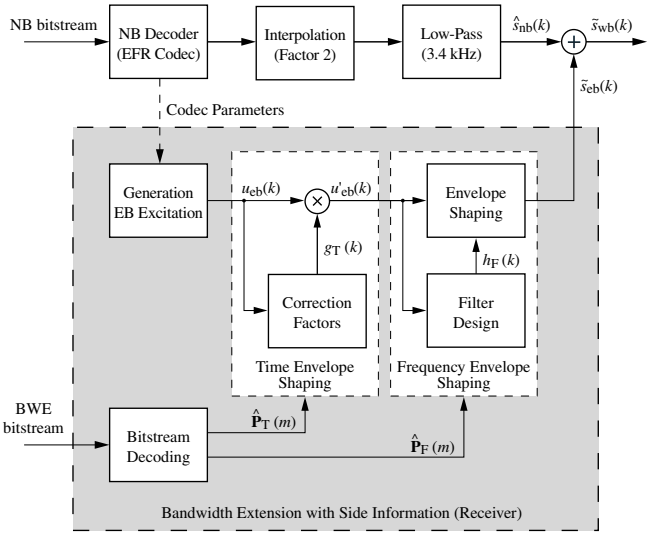

Fig. 2. BWE receiver with time and frequency envelope shaping of an artificially generated excitation signal.

3.1. Generation of the Excitation Signal

The excitation signal $u_{\mathrm{eb}}(k)$ serves as input to the time and frequency envelope shaping blocks. Since these two blocks can only reconstruct the envelope characteristics of the EB signal components, the minimum requirement to $u_{\mathrm{eb}}(k)$ is that a spectrally flat signal with a correct harmonic structure should be produced.

In the proposed algorithm the excitation signal is estimated using a number of parameters taken from the narrowband decoder, see Fig. 3. In particular, we use the fixed and adaptive codebook parameters of the CELP narrowband codec. The principle of the excitation generation is to run the LTP synthesis filter at an increased sampling frequency of 48 kHz, thereby producing harmonics of the pitch frequency also in the EB frequency range. The input to the LTP synthesis filter is obtained by inserting five zeros between every other sample of the fixed codebook contributions in the EFR decoder, thereby producing (mirrored) signal components up to a frequency of 24 kHz. Since the LTP filter is now operating at 48 kHz the utilized pitch lag has to be six times the pitch lag of the narrowband codec. Thereby, the system can make full use of a fractional sample resolution in the EFR pitch lag information. The output of the LTP synthesis filter is band-pass filtered and decimated. The excitation signal $u_{\mathrm{eb}}(k)$ has a sampling rate of 16 kHz.

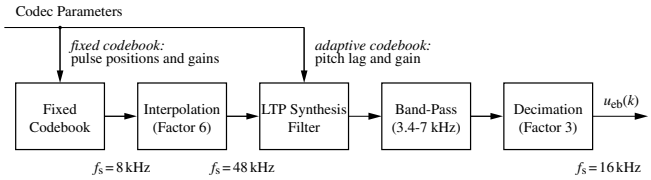

Fig. 3. Generation of the excitation signal. Note that the interpolation and decimation blocks do *not* include low-pass filtering.

3.2. Time Envelope Shaping

The time envelope of the EB signal components is shaped by a scalar correction factor g_{T} that is multiplied to each sample of the excitation signal, $u'_{\mathrm{eb}}(k) = g_{\mathrm{T}}(k)\, u_{\mathrm{eb}}(k)$. Recall that multiplication in time domain corresponds to a convolution operation in frequency domain. Since the frequency representation of $u_{\mathrm{eb}}(k)$ shall in principle not be altered by the time envelope shaping block, the gain function has to have strict lowpass frequency characteristics.

To determine the gain function $g_{\mathrm{T}}(k)$, the excitation signal $u_{\mathrm{eb}}(k)$ is segmented and analyzed in the same manner as described in Sec. 2 for the time envelope extraction from the original EB signal in the BWE encoder. The ratio between the decoded target power $\hat{T}_{m,n}$ and the analysis result $T^{\mathrm{R}}_{m,n}$ leads to the desired gain factor for the n-th subframe of the m-th signal segment

$$g'_{\mathrm{T}}(m,n) = \sqrt{\frac{\hat{T}_{m,n}}{T^{\mathrm{R}}_{m,n}}}. \tag{4}$$

The final gain function is determined by placing single pulses, weighted by the respective gain factors from (4), into the middle of each subframe. Subsequently this sparse sequence of pulses is filtered to obtain $g_{\mathrm{T}}(k)$, using the Hann window $w_{\mathrm{T}}(k)$ (the same as used for time envelope extraction, see Sec. 2.1) as filter coefficients. Thereby, the gain function exhibits the required low-pass character.

3.3. Frequency Envelope Shaping

After time envelope shaping of the excitation signal, resulting in the signal $u'_{\mathrm{eb}}(k)$, the next step is shaping of the frequency envelope. The concept is similar to that already described for time envelope shaping: the input signal $u_{\mathrm{eb}}(k)$ is analyzed to obtain the frequency envelope information $F^{\mathrm{R}}_{m,l,\nu}$ (see Sec. 2.2). This observed envelope is compared to the envelope $\hat{F}_{m,l,\nu}$ as decoded from the BWE bitstream. Thus, for each sub-band (index ν) of the frequency envelope representation a correction gain is determined

$$G_{\mathrm{F}}(m,l,\nu) = \sqrt{\frac{\hat{F}_{m,l,\nu}}{F^{\mathrm{R}}_{m,l,\nu}}}. \tag{5}$$

Using these sub-band gains, for each subframe a set of filter coefficients $h_{\mathrm{F}}(k)$ is computed which is applied in a time domain FIR filter to achieve the desired frequency envelope shaping. That is, the output of the frequency envelope shaping block is obtained by

$$\tilde{s}_{\mathrm{eb}}(k) = \sum_{\kappa} u'_{\mathrm{eb}}(k-\kappa)\, h_{\mathrm{F}}(\kappa). \tag{6}$$

The filter output $\tilde{s}_{\mathrm{eb}}(k)$ can be regarded as an estimate of the EB signal. The wideband speech signal $\tilde{s}_{\mathrm{wb}}(k)$ is determined by adding the estimated EB signal and the decoded NB signal $\hat{s}_{\mathrm{nb}}(k)$.

The coefficients $h_{\mathrm{F}}(k)$ have to be determined anew for each subframe according to the respective subband gain factors $G_{\mathrm{F}}(m,l,\nu)$ from (6). This is accomplished by a weighted sum of prototype band-pass filters $h_{\mathrm{F}}^{(\nu)}(k)$, i.e.,

$$h_{\mathrm{F}}(k) = \sum_{\nu=0}^{N_{\mathrm{sb}}-1} G_{\mathrm{F}}(m,l,\nu)\, h_{\mathrm{F}}^{(\nu)}(k). \tag{7}$$

All of the sub-band prototype filters are defined by modulation of a low-pass prototype filter $h_{\mathrm{lp}}(k)$ which is determined by a Kaiser window $w(k)$ of length 65 with $\beta = 5$

$$h_{\mathrm{lp}}(k) = \frac{10^{\frac{-1.75}{20}}}{\sum_{\kappa=0}^{64} w(\kappa)} \cdot w(k) \tag{8}$$

The ν-th sub-band prototype filter is then defined by

$$h_{\mathrm{F}}^{(\nu)}(k) = 2\, h_{\mathrm{lp}}(k) \cdot \cos\left(\frac{3450 + 375\,(\nu - 0.5)}{8000}\, 2\pi k\right). \tag{9}$$

All of the filters have linear phase with a delay of 32 samples (2 ms).

4. PERFORMANCE EVALUATION

To investigate the performance of our new approach, we have implemented the BWE algorithm in combination with the GSM EFR speech codec [1] (bit rate 12.2 kbit/s). We have then conducted listening tests to compare the subjective quality of our approach with that of well-known speech codecs. These reference codecs were the AMR-WB codec at a bit rate of 12.65 kbit/s, the G.722 codec at 64 kbit/s, and the EFR codec at 12.2 kbit/s without any bandwidth extension add-on. We have tested our algorithm with three different bit rates for the BWE side information bitstream, namely 0.3 kbit/s, 0.6 kbit/s, and 1.5 kbit/s. We have used a MUSHRA-style test according to ITU-R BS.1534. The test was conducted with clean English speech signals of male and female speakers. Ten experienced listeners participated in the test.

The test results are illustrated in Fig. 4. For each codec the mean MUSHRA score and the 95% confidence intervals are given. We can observe that the subjective quality of the BWE scheme is always better than that of the "raw" narrowband EFR codec. Even as little as 300 bit/s of additional side information produce a consistently improved speech quality. With 1.5 kbit/s of side information (gross bit rate of 13.7 kbit/s) the BWE approach has about the same subjective quality as the AMR-WB codec with 12.65 kbit/s.

We have performed further informal listening tests to evaluate the performance in background noise conditions and for music signals. While the proposed approach does not produce significant artifacts with background noise, it is not well suited for coding of music signals. This is partly due to the limited suitability of the EFR codec for music signals, but also stems from the fact that the BWE design is focused on the properties of speech signals only, mainly with respect to the generation of the excitation signal.

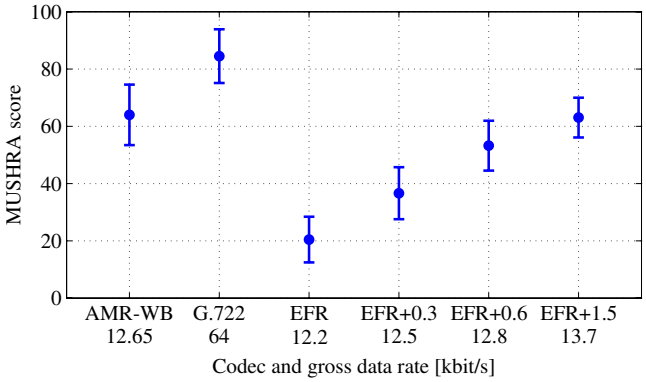

Fig. 4. Results of the MUSHRA-style subjective listening test.

5. CONCLUSION

We have proposed a new embedded wideband coding concept based on bandwidth extension with side information on top of a standard narrowband speech codec. The major differences to previous approaches are that we perform frequency envelope shaping of the EB signal components not with LPC techniques but rather by a time-variant linear-phase FIR filter, and that the transmitted time envelope has a better time resolution. Despite its conceptual simplicity the new algorithm results in a natural sounding speech quality.

Although presented and evaluated here for the GSM EFR codec, the scheme can be applied to a wide range of existing speech codecs. Together with the bit rate scalability of the proposed PCA based quantization scheme, this gives a very high flexibility to tune the algorithm. Therefore, and since the BWE "add-on" of the NB codec produces significant quality improvements for comparatively low additional bit rates, the proposed concept is a candidate for a large number of interesting applications.

6. REFERENCES

[1] ETSI Recommendation GSM 06.60, "Enhanced full rate (EFR) speech transcoding," version 8.0.1, release 1999, Nov. 2000.

[2] 3GPP TS 26.190, "AMR wideband speech codec; transcoding functions," Dec. 2001.

[3] P. Jax and P. Vary, "On artificial bandwidth extension of telephone speech," *Signal Processing*, vol. 83, no. 8, pp. 1707–1719, Aug. 2003.

[4] P. Jax, "Bandwidth extension for speech," in *Audio Bandwidth Extension*, E. Larsen and R. M. Aarts, Eds., chapter 6, pp. 171–236. Wiley and Sons, Nov. 2004.

[5] 3GPP TS 26.290, "Extended AMR wideband codec; transcoding functions," Sept. 2004.

[6] 3GPP TS 26.404, "Enhanced aacPlus general audio codec; encoder specification; spectral band replication (SBR) part," Sept. 2004.

[7] J. W. Paulus and J. Schnitzler, "16 kbit/s wideband speech coding based on unequal subbands," in *Proc. of ICASSP*, Atlanta, GA, USA, May 1996, pp. 255–258.

[8] J. Schnitzler and P. Vary, "Trends and perspectives in wideband speech coding," *Signal Processing*, vol. 80, no. 11, pp. 2267–2281, Nov. 2000.

[9] R. Taori, R. J. Sluijter, and A. J. Gerrits, "Hi-BIN: An alternative approach to wideband speech coding," in *Proc. of ICASSP*, Istanbul, Turkey, June 2000, vol. 2, pp. 1157–1160.

[10] A. McCree, "A 14 kb/s wideband speech coder with a parametric high-band model," in *Proc. of ICASSP*, Istanbul, Turkey, June 2000, vol. 2, pp. 1153–1156.

[11] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Morgan Kaufmann, Academic Press, 2nd edition, 1990.

[12] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Transactions on Communications*, vol. 28, no. 1, pp. 84–95, Jan. 1980.