BACKWARDS COMPATIBLE WIDEBAND TELEPHONY IN MOBILE NETWORKS: CELP WATERMARKING AND BANDWIDTH EXTENSION

Bernd Geiser and Peter Vary

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

ABSTRACT

We consider the problem of transmitting a *wideband speech signal* with a cut-off frequency of $f_c = 7\,\text{kHz}$ over a standardized *narrowband* ($f_c = 3.4\,\text{kHz}$) communication link in a *backwards compatible* manner. In a previous contribution [1] we have shown that backwards compatibility can be achieved by using *digital watermarking*: We embedded compact side information about the missing high frequency band ($3.4-7\,\text{kHz}$) into the narrowband speech signal. Here, we present a related system which is especially tailored to state-of-the-art narrowband speech communication networks such as GSM or UMTS. Therefore, we propose an improved *low bit rate bandwidth extension algorithm* along with a *robust watermarking scheme* for CELP-type speech codecs. The practical relevance of our system is shown by speech quality evaluations and by link-level simulations for the "Enhanced Full Rate Traffic Channel" (TCH/EFS) of the GSM cellular communication system.

Index Terms— Speech coding, speech codecs, land mobile radio cellular systems

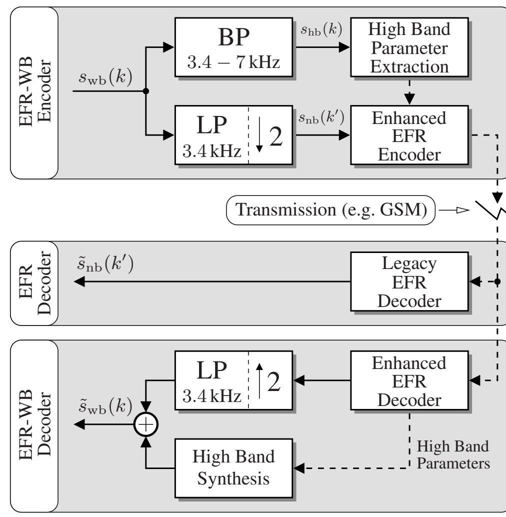

Fig. 1. Transmission system for backwards compatible wideband telephony using the GSM EFR narrowband codec [5].

1. INTRODUCTION

The transmission of wideband speech signals with a cut-off frequency of at least 7 kHz is a highly desirable feature for future speech communication networks, offering significantly increased subjective speech quality and intelligibility. Suitable wideband speech codecs — such as the AMR-WB [2] — have been developed in the past. However, the requirement of *backwards compatibility* with existing narrowband equipment turned out to be an obstacle for the introduction of wideband speech coding in today's communication networks.

A promising possibility to resolve this dilemma is the deployment of *artificial bandwidth extension* (BWE), i.e., pure *estimation* of missing frequency components from the narrowband signal. Such techniques might, as anticipated in [3], be able to speed up the narrow- to wideband change-over. Though, the inherently limited performance of stand-alone BWE algorithms [4] is not sufficient for the regeneration of *high quality* wideband speech signals. Therefore, we introduced the idea of embedding a limited amount of *side information* about the missing frequency band into the narrowband speech signal via *digital watermarking* [1]. Hence, using the decoded side information, the respective receiver side BWE algorithm can produce a wideband signal of higher quality than the stand-alone BWE approach. Thereby, backwards compatibility w.r.t. legacy narrowband terminals — and the network as well — is maintained.

In [1] the watermark is embedded into the narrowband speech signal *prior* to coding, i.e., the system does not take the implemented speech codec into account and the *coding distortion* directly impacts the embedded watermark. Thus, when using state-of-the-art CELP codecs like the GSM "Enhanced Full Rate" (EFR) codec [5], the watermark bits are severely disturbed even before they are transmitted. Consequently, as the watermark bit error rate may be very high, powerful error concealment mechanisms are indispensible[1]. The solution to enable *robust* watermarking of CELP coded signals is to integrate the information embedding *into* the speech coder. A related approach has recently been taken for the standard GSM "Full Rate" RELP codec [6]. This contribution presents a suitable solution for state-of-the-art CELP codecs (Sec. 3).

Apart from a robust watermarking scheme, the application of backwards compatible wideband speech coding also requires a *low bit rate BWE algorithm*. Several proposals that offer the relevant bit rates, i.e., below 1 kbit/s, have been made in the literature, e.g. [7, 8]. Here, we use an improved version of the algorithm from [8] and apply it to the GSM EFR narrowband codec [5] (Sec. 2).

The respective backwards compatible transmission system, which is illustrated in Fig. 1, is analyzed and evaluated in Sec. 4. We present speech quality assessment results for narrowband and wideband decoders along with results of link-level simulations for the "Enhanced Full Rate Traffic Channel" (TCH/EFS) of the GSM cellular communication system.

[1] In [1] the error concealment has been achieved by integrating the watermark decoding and a stand-alone BWE into a uniform estimation framework.

2. LOW BIT RATE BANDWIDTH EXTENSION

This section describes a low bit rate BWE algorithm that is based on on our BWE module from [8] which has been adopted as a part of ITU-T Rec. G.729.1 [9]. A rather low bit rate of 400 bit/s is obtained by sacrificing the enhanced temporal resolution of [8] and by using a predictive quantization scheme which reuses information from the low band signal.

2.1. Band Split and Recombination

The "high band parameter extraction" shown in Fig. 1 operates on 20 ms frames of a band-pass filtered signal $s_{\mathrm{hb}}(k)$, sampled at $f_{\mathrm{s}} = 16$ kHz, with $k \in \{0, \dots, L-1\}$, $L = 320$. In parallel, the low-pass filtered and downsampled signal $s_{\mathrm{nb}}(k')$, sampled at $f'_{\mathrm{s}} = 8$ kHz, with $k' \in \{0, \dots, L'-1\}$, $L' = 160$, is encoded by the narrowband EFR encoder. In the *wideband decoder*, the low band signal $\tilde{s}_{\mathrm{nb}}(k')$ is synthesized by the embedded EFR decoder, and an interpolated version is added to the synthetically generated high band signal $\tilde{s}_{\mathrm{hb}}(k)$. The wideband output is $\tilde{s}_{\mathrm{wb}}(k)$.

2.2. Parameter Extraction

The parameter set of our low bit rate BWE comprises a (logarithmic) *high band gain* g_{hb} and a *spectral envelope*, i.e., (logarithmic) DFT-domain *sub-band energies* $F_{\mathrm{hb}}(\nu)$. The gain g_{hb} is given by

$$g_{\mathrm{hb}} = \frac{1}{2}\,\mathrm{ld}\left(\frac{1}{L}\sum_{k=0}^{L-1} s_{\mathrm{hb}}^2(k)\right). \qquad (1)$$

The spectral envelope computation requires the DFT of $s_{\mathrm{hb}}(k)$, i.e., $S_{\mathrm{hb}}(\mu) = \sum_{k=0}^{L_{\mathrm{F}}-1} w_{\mathrm{F}}(k) \cdot s_{\mathrm{hb}}(k+L-L_{\mathrm{F}}) \cdot \exp\left(-j\frac{2\pi k\mu}{L_{\mathrm{F}}}\right)$, where the DFT window length is $L_{\mathrm{F}} = 256$. We use a slightly asymmetric Hann window $w_{\mathrm{F}}(k)$ with a look-ahead of 32 and a look-back of 64 taps. Note that, due to the offset $(L - L_{\mathrm{F}})$, the spectrum $S_{\mathrm{hb}}(\mu)$ is representative for the *second* 10 ms sub-frame within the current frame. The parameters for the *first* sub-frame are interpolated by the decoder. Now, based on the power spectrum $|S_{\mathrm{hb}}(\mu)|^2$, the energies for $N_{\mathrm{sb}} = 10$ partially overlapping sub-bands are computed:

$$F_{\mathrm{hb}}(\nu) = \frac{1}{2}\mathrm{ld}\sum_{\mu=0}^{L_{\mathrm{W}}-1} W_{\mathrm{F}}(\mu)\,|S_{\mathrm{hb}}(\mu + \mu_0 + \nu(L_{\mathrm{W}} - L_{\mathrm{o}}))|^2 \qquad (2)$$

with $\nu \in \{0, \dots, N_{\mathrm{sb}} - 1\}$. The frequency domain window $W_{\mathrm{F}}(\mu)$ is a flat-top Hann window of length $L_{\mathrm{W}} = 9$. The spectral overlap is $L_{\mathrm{o}} = 3$ and the high frequency band starts at DFT bin $\mu_0 = 54$.

2.3. Quantization

In order to achieve a good performance for the quantization of g_{hb} and $\mathbf{F}_{\mathrm{hb}} \doteq (F_{\mathrm{hb}}(0), \dots, F_{\mathrm{hb}}(N_{\mathrm{sb}}-1))^{\mathrm{T}}$ at the target bit rate of 400 bit/s, we employ a predictive scheme which reuses information about the low band signal. For the quantization of g_{hb}, this information is gathered in the *feature vector* $\mathbf{x}_{\mathrm{f},1} \doteq (\hat{g}_{\mathrm{lb,old}}, \hat{\mathbf{q}})^{\mathrm{T}}$, where $\hat{g}_{\mathrm{lb,old}}$ is, analogous to (1), the gain of the *coded* low band signal from the *previous*[2] speech frame. The vector $\hat{\mathbf{q}}$ comprises the quantized LSP coefficient set from the EFR narrowband coder. Based on $\mathbf{x}_{\mathrm{f},1}$, we compute a *linear estimate* (see box below) of the high band gain: $\tilde{g}_{\mathrm{hb}} = \mathbf{H}_1^{\mathrm{T}} \cdot \mathbf{x}_{\mathrm{f},1}$. The residual estimation error is quantized afterwards using a Lloyd-Max quantizer with 3 bit: $\Delta\hat{g}_{\mathrm{hb}} = \mathcal{Q}_1(g_{\mathrm{hb}} - \tilde{g}_{\mathrm{hb}})$, i.e., the quantized high band gain is given by: $\hat{g}_{\mathrm{hb}} = \Delta\hat{g}_{\mathrm{hb}} + \tilde{g}_{\mathrm{hb}}$. A similar procedure applies for the spectral envelope parameter set $\mathbf{F}_{\mathrm{hb}}$. The respective feature vector comprises the quantized high band gain: $\mathbf{x}_{\mathrm{f},2} \doteq (\hat{g}_{\mathrm{hb}}, \hat{\mathbf{q}})^{\mathrm{T}}$. Again, an estimate is computed: $\tilde{\mathbf{F}}_{\mathrm{hb}} = \mathbf{H}_2^{\mathrm{T}} \cdot \mathbf{x}_{\mathrm{f},2}$, and the estimation error is quantized: $\Delta\hat{\mathbf{F}}_{\mathrm{hb}} = \mathcal{Q}_2(\mathbf{F}_{\mathrm{hb}} - \tilde{\mathbf{F}}_{\mathrm{hb}})$. The vector-quantizer $\mathcal{Q}_2$ uses 5 bit. The quantized spectral envelope is $\hat{\mathbf{F}}_{\mathrm{hb}} = \Delta\hat{\mathbf{F}}_{\mathrm{hb}} + \tilde{\mathbf{F}}_{\mathrm{hb}}$. In total, this scheme yields a bit rate of $(3 + 5)$ bit/20 ms = 400 bit/s.

[2] This is due to the fact that the gain of the *current* frame can not be computed as yet, because the high band information still needs to be *embedded* into the narrowband signal. Though, it is possible to include the LSP coefficients from the *current* frame in $\mathbf{x}_{\mathrm{f},1}$ which will become evident in Sec. 3.

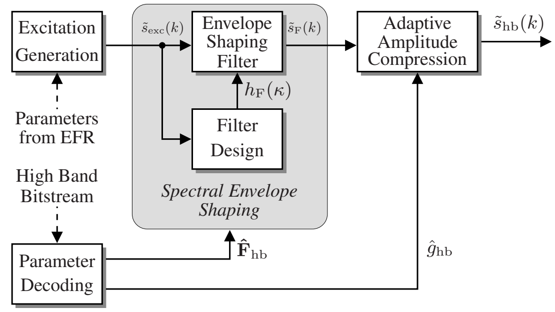

Fig. 2. Signal flow chart for high band synthesis.

Linear Estimation [10]
Consider the linear estimation task $\tilde{\mathbf{y}} = \mathbf{H}^{\mathrm{T}} \cdot \mathbf{x}_{\mathrm{f}}$ with the feature vector $\mathbf{x}_{\mathrm{f}}$ and the estimation matrix $\mathbf{H}$. Given a training set of feature vectors, collected in the *rows* of the matrix $\mathbf{X}$ and the corresponding matrix of training data $\mathbf{Y}$ consisting of representative vectors of the quantity to be estimated, the optimum matrix $\mathbf{H}$ is given by the pseudo-inverse of $\mathbf{X}$ multiplied with $\mathbf{Y}$:

$$\mathbf{H} = \mathbf{X}^{+} \cdot \mathbf{Y} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}} \cdot \mathbf{Y}.$$

2.4. Synthesis

The "high band synthesis" block from Fig. 1 produces the signal $\tilde{s}_{\mathrm{hb}}(k)$ using the BWE scheme shown in Fig. 2. Its individual components are briefly sketched below.

Excitation Generation — This module has been adapted from [9] with modifications to produce an excitation signal $\tilde{s}_{\mathrm{exc}}(k)$ in the desired frequency range. The algorithm places pulses with a *spacing* that corresponds to the *integer* part of a post-processed EFR pitch lag. The pulse *shape* is selected according to the *fractional* pitch lag, thus preserving the sub-sample resolution. Afterwards, the resulting signal is mixed with noise and band-pass filtered. The *weighting* of tonal and noisy contributions is derived from the energy ratio of the LTP and fixed codebook contributions of the EFR narrowband coder.

Spectral Envelope Shaping — First, the *two* 10 ms sub-frames of $\tilde{s}_{\mathrm{exc}}(k)$ are analyzed as described by (2) and the resulting values $F_{\mathrm{exc}}^{(1,2)}(\nu)$ are used to compute correction gains per sub-band: $G^{(1)}(\nu) = 2^{F_{\mathrm{hb}}^{\mathrm{interp}}(\nu) - \hat{F}_{\mathrm{exc}}^{(1)}(\nu)}$ and $G^{(2)}(\nu) = 2^{\hat{F}_{\mathrm{hb}}(\nu) - F_{\mathrm{exc}}^{(2)}(\nu)}$, where, for the *first* 10 ms sub-frame, $F_{\mathrm{hb}}^{\mathrm{interp}}(\nu)$ is the linear interpolation between the two adjacent envelope parameters $\hat{F}_{\mathrm{hb}}(\nu)$ and $\hat{F}_{\mathrm{hb,old}}(\nu)$. As in [8], the gains $G^{(1,2)}(\nu)$ control the channels of a *filter-bank equalizer* with band-pass impulse responses $h_\nu(\kappa)$ that match the sub-band division which is implicitly given by (2):

$$h_{\mathrm{F}}^{(1,2)}(\kappa) = \sum_{\nu=0}^{N_{\mathrm{sb}}-1} G^{(1,2)}(\nu) \cdot h_\nu(\kappa). \qquad (3)$$

The resulting linear phase FIR filters $h_{\mathrm{F}}^{(1,2)}(\kappa)$ of length $L_{h_{\mathrm{F}}} = 65$ are then applied to the respective sub-frames of $\tilde{s}_{\mathrm{exc}}(k)$ to give the spectrally shaped signal $\tilde{s}_{\mathrm{F}}(k)$. The filter delay is 2 ms.

Adaptive Amplitude Compression — Here, large deviations of $|\tilde{s}_{\mathrm{F}}(k)|$ from the (linear) high band gain $2^{\hat{g}_{\mathrm{hb}}}$ are attenuated (cf. [9]).

3. ROBUST WATERMARKING FOR CELP CODECS

When applying conventional watermarking techniques, such as spread-spectrum or quantization based methods [11], to a speech signal which is then transcoded by a state-of-the-art CELP codec, it turns out that the watermark transmission is not very reliable if the watermark is supposed to be imperceptible and data rates of several 100 bit/s are required. The reason for this is that the CELP transcoding process itself disturbs the embedded message. This becomes obvious by recognizing that watermarks, to be imperceptible, are usually embedded into *less relevant components* of the speech signal. Unfortunately, such speech components are also very likely to be coarsely quantized or omitted by the speech coder.

To circumvent the disadvantages of time or transform domain watermarking, we propose the idea to *integrate* the information embedding *into the speech encoder*. A respective embedding rule can take the coder's specific characteristics into account and should be designed such that the information embedding does not (or only insignificantly) degrade the narrowband speech quality and that the embedded message can be reliably reconstructed at least as long as the codec bits are not disturbed. Our particular application imposes two further constraints: The information embedding has to be performed per frame, i.e., we disallow any additional algorithmic delay and the embedding rate (e.g., 400 bit/s) may not vary over time.

Reconsidering the parameters which are typically transmitted by a CELP coder, it is rather obvious that any modification regarding spectral envelope (LSFs), pitch lag, or gain factors is likely to be perceivable. Hence, watermark embedding should focus on the bits related to the coding of the innovative codevector. Though, a direct "abuse" of these bits for watermarking is not satisfactory either.

3.1. Concept

An attractive solution is found in the *binning scheme* [11] which is briefly introduced in the box below. For the case of a fixed codebook search in CELP coders, the cost function to be minimized is

$$\chi(\mathbf{c}) = ||\mathbf{v}||^2 - \frac{\left(\mathbf{v}^{\mathrm{T}}\mathbf{H}\mathbf{c}\right)^2}{||\mathbf{H}\mathbf{c}||^2} \tag{4}$$

with the target vector $\mathbf{v}$, the perceptually weighted filter matrix $\mathbf{H}$, and the innovative codebook vectors $\mathbf{c} \in \mathcal{C}$. In practical codebook search algorithms the *second* term of $\chi(\mathbf{c})$ is *maximized*. *Information embedding* of the message m' into the selected codevector $\hat{\mathbf{c}}$ is, according to the "Binning Scheme", done by minimizing $\chi(\mathbf{c})$ over the *subset* $\mathcal{C}_{m'}$ of the fixed codebook $\mathcal{C}$. This way, codebook search and information embedding are performed *jointly in the analysis-by-synthesis loop*.

Furthermore, as the codebook $\mathcal{C}$ in practical *algebraic* CELP coders (ACELP) is typically very large, only a small heuristically selected subset $\mathcal{C}' \subset \mathcal{C}$ is examined during the codebook search, i.e., codebook search in practical CELP coders is by far *non-exhaustive*. This fact can be exploited in the design of codebook partitionings for information embedding via binning. Specifically, the sub-codebooks $\mathcal{C}_m \subset \mathcal{C}$ may be chosen to have a cardinality equal to (or even larger than) the cardinality of the originally examined subset $\mathcal{C}'$. Thus, for a good codebook partitioning (usually achieved with *nested* sub-codebooks), the search procedure is likely to find good codevectors in the (extended) sub-codebooks $\mathcal{C}_m$. This leads to a pretty low speech quality loss due to information embedding at a limited expense of additional computational complexity.

Information Embedding via "Binning Schemes"
Consider a search procedure for the codebook $\mathcal{C}$ aiming at the minimization of the cost function $\chi(\mathbf{c})$ with $\mathbf{c} \in \mathcal{C}$, i.e., $\hat{\mathbf{c}} = \arg\min_{\mathbf{c}\in\mathcal{C}} \chi(\mathbf{c})$. For the *embedding* of N bits into the selected codevector $\hat{\mathbf{c}}$, the codebook $\mathcal{C}$ is *partitioned* into $M = 2^N$ *disjoint sub-codebooks* $\mathcal{C}_m$ with $m \in \{0, \ldots, M-1\}$ such that $\bigcup_{m=0}^{M-1} \mathcal{C}_m \subseteq \mathcal{C}$. Codebook search *with information-embedding* is defined as $\hat{\mathbf{c}} = \arg\min_{\mathbf{c}\in\mathcal{C}_{m'}} \chi(\mathbf{c})$, where m' is the *message* to be embedded. The *decoder* only needs to identify the sub-codebook that contains $\hat{\mathbf{c}}$, i.e., m' is implicitly given by $\hat{\mathbf{c}} \cap \mathcal{C}_{m'} = \hat{\mathbf{c}}$.

3.2. Application to the GSM EFR Codec

We applied the concept outlined in Sec. 3.1 to the GSM Enhanced Full Rate Codec [5]. This ACELP codec uses a fixed codebook with an algebraic structure. Per 5 ms sub-frame (40 samples), the codevectors $\mathbf{c}$ contain 10 signed unit pulses and zeros otherwise, i.e., $c_i \in \{-1, 0, +1\}$ with $i \in \{0, \ldots, 39\}$, where 2 pulses are placed on each of the 5 tracks (interleaved sub-grids). Thereby, the *pulse signs* are pre-computed "out of the loop" and coded with 5 bit per sub-frame. Therefore, we perform information embedding into the $10 \cdot 3\,\mathrm{bit} = 30\,\mathrm{bit}$ per sub-frame ($30\,\mathrm{bit}/5\,\mathrm{ms} = 6\,\mathrm{kbit/s}$) which are used for the closed loop encoding of the *pulse positions*.

Codebook Partitioning — Let $(m)_i$, $i \in \{0, \ldots, N-1\}$, denote the individual bits of the message m to be embedded into one sub-frame. Further, let $(c)_j$, $j \in \{0, \ldots, 29\}$, denote the codeword bits that correspond to the examined codevectors (pulse positions). The modified analysis-by-synthesis loop *only* examines codevectors which result in codewords with bits $(c)_j$ that match a certain parity condition:

$$(m)_i = \bigoplus_{j\in\mathbb{A}_i} (c)_j \quad \forall\, i \in \{0, \ldots, N-1\}\,, \tag{5}$$

where $\oplus$ denotes "sum modulo 2" and the (unique but not necessarily disjoint) sets $\mathbb{A}_i \subset \{0, \ldots, 29\}$ specify the bits to be considered for the parity computation. For the embedding of, e.g., $N = 2\,\mathrm{bit}$ per sub-frame (a total of 400 bit/s) it turned out to be advantageous to choose the set $\mathbb{A}_0$ as the $\alpha = 5$ least significant bits (LSBs) of the *first* pulses of each track and $\mathbb{A}_1$ as the LSBs of the respective *second* pulses. This choice guarantees a sufficiently nested codebook partitioning, hence, low embedding distortion.

In addition, the EFR codebook search algorithm has been modified to examine more pulse position permutations than specified by the standard. Together with (5), this ensures that the narrowband signal with 400 bit/s of embedded information suffers *no quality loss* as compared to the standard EFR implementation. For higher embedding rates a certain degradation can be observed (see Sec. 4).

Robustness to Transmission Errors — The proposed embedding scheme allows a *perfect reconstruction* of the embedded message as long as the codec bits are not disturbed. However, if these bits are transmitted over a noisy channel, there will also be a certain error probability $p_{\mathrm{e,wm}}$ for the *embedded bits*. Suppose we have embedded *one* message bit into α transmitted codec bits which have an average bit error probability of p_{e} (after channel decoding). From this, albeit the unequal error protection, $p_{\mathrm{e,wm}}$ can be approximated:

$$p_{\mathrm{e,wm}}(\alpha, p_{\mathrm{e}}) \approx \sum_{0\leq n<\alpha/2} \binom{\alpha}{2n+1} \cdot p_{\mathrm{e}}^{2n+1} \cdot (1-p_{\mathrm{e}})^{\alpha-2n-1}. \tag{6}$$

Hence, the more bits α are considered for information embedding, the less robust is the watermark transmission. Thus, the sets $\mathbb{A}_i$ in (5) should not contain too many elements.

Moreover, the *gray coding* of the pulse position codewords $(c)_j$ in the EFR coder [5] has to be considered. For reasons of increased robustness it is beneficial to compute the parity condition (5) over the gray coded bits $(\mathcal{G}(c))_j$ instead of $(c)_j$.

4. EVALUATION AND TEST RESULTS

Narrowband Speech Quality — Here, the speech quality is rated for the case of our *enhanced encoder* in combination with a *"legacy" decoder* which decodes a bitstream with embedded information of rate r. Table 1 lists objective narrowband MOS scores (mapped PESQ, P.862.2 NB mode, obtained from 88 s of English speech) for various watermark embedding rates. Also, for reference, the standard EFR codec is evaluated. Note that the MOS scores for $r = 0$ and $r = 200$ bit/s are higher than for the standard EFR. This is due to the extended codebook search procedure. We note that potentially even more quality could be gained, especially at high embedding rates, by further extending the codebook search.

Table 1. Narrowband speech quality for various embedding rates r. For reference, the last column represents the *standard* EFR codec.

r[bit/s]	0	200	**400**	600	800	1000	EFR
MOS-LQO	4.19	4.17	**4.16**	4.07	3.86	3.66	4.16

Wideband Speech Quality — Table 2 presents objective wideband MOS scores (WB-PESQ measurements for 88 s of English speech). Our "EFR-WB" coder, using 400 bit/s of hidden side-information for high band synthesis, is compared with G.729.1@14 kbit/s [9], which uses a very similar BWE technique, and with the AMR-WB codec [2]. Actually, the EFR-WB performs very well in this test which is, of course, also the merit of the EFR low band coder.

Table 2. Wideband speech quality evaluations.

Codec	G.729.1	AMR-WB	AMR-WB	EFR-WB
Bit Rate	14 kbit/s	8.85 kbit/s	12.65 kbit/s	12.2 kbit/s
WB MOS-LQO	3.60	3.05	3.46	3.46

GSM Link-Level Simulations — The simulations for the GSM EFR traffic channel (TCH/EFS) have been conducted using the TU50 channel profile (typical urban, 50 km/h) [12]. Fig. 3 shows the results of narrow- and wideband speech quality assessments and bit error rate evaluations for the *codec bits* as well as for the *embedded bits*. For our configuration ($\alpha = 5$ in Eq. (6)), the wideband decoder's speech quality is rated superior for SNR values above 10 dB.

The "decoder with *error concealment*" is merely a feasibility study since the bit error *locations* are assumed to be known. A *practical* error concealment scheme could, e.g., slightly increase the embedding rate and add some CRC bits to the hidden information.

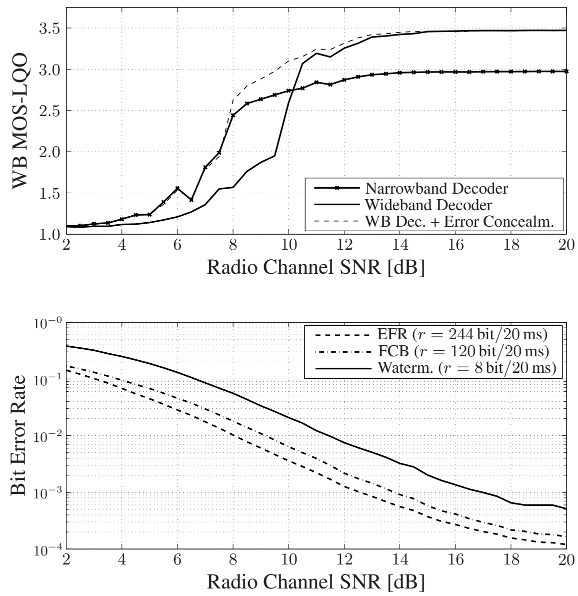

Fig. 3. Results of the GSM link-level simulations for the TU50 channel profile: Objective *wideband speech quality* (WB-PESQ) for narrow-/wideband decoders and *bit error rates* for all EFR bits, for the fixed codebook bits, and for the watermark bits.

5. CONCLUSION AND OUTLOOK

In this contribution we have presented a "backwards compatible wideband codec" based on the GSM EFR narrowband coder and on bandwidth extension with 400 bit/s of side information which is *embedded* into the EFR bitstream. The information embedding is performed *jointly* with the CELP analysis-by-synthesis codebook search. The obtained speech quality, both for the narrow- and the wideband case, and the link-level simulations demonstrate the practical relevance of our proposal, at least for *tandem-free* operation in mobile communication networks. The required additional algorithmic delay and complexity figures are expected to be moderate, and a preliminary informal listening test indicates that the subjective (narrowband and wideband) speech quality is almost as good as the objective scores suggest.

The proposed "EFR-Wideband" codec does not yet provide a *wideband VAD/DTX/CNG* scheme. Also, as discussed above, a proper *error concealment* mechanism would be desirable. Finally, appropriate (hidden) *signaling* for wideband operation is required.

6. REFERENCES

[1] B. Geiser, P. Jax, and P. Vary, "Artificial bandwidth extension of speech supported by watermark transmitted side information," in *Proc. of INTERSPEECH*, Lisbon, Portugal, Sept. 2005, pp. 1497–1500.

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

[3] P. Jax and P. Vary, "Bandwidth extension of speech signals: A catalyst for the introduction of wideband speech coding?" *IEEE Communications Magazine*, vol. 44, no. 5, pp. 106–111, May 2006.

[4] ——, "An upper bound on the quality of artificial bandwidth extension of narrowband speech signals," in *Proc. of ICASSP*, vol. 1, Orlando, FL, USA, May 2002, pp. 237–240.

[5] ETSI Rec. GSM 06.60, "Digital Cellular Telecommunication System (Phase 2+); Enhanced Full Rate (EFR) speech transcoding," version 8.0.1, release 1999, Nov. 2000.

[6] N. Chétry and M. Davies, "Embedding side information into a speech codec residual," in *Proc. of EUSIPCO*, Florence, Italy, Sept. 2006.

[7] J.-M. Valin and R. Lefebvre, "Bandwidth extension of narrowband speech for low bit-rate wideband coding," in *IEEE Speech Coding Worksh.*, Delavan, WI, USA, Sept. 2000, pp. 130–132.

[8] P. Jax, B. Geiser, S. Schandl, H. Taddei, and P. Vary, "An embedded scalable wideband codec based on the GSM EFR codec," in *Proc. of ICASSP*, vol. 1, Toulouse, France, May 2006, pp. 5–8.

[9] ITU-T Rec. G.729.1, "G.729 based embedded variable bit-rate coder: An 8-32 kbit/s scalable wideband coder bitstream interoperable with G.729," 2006.

[10] S. M. Kay, *Fundamentals of Statistical Signal Processing — Estimation Theory*. Upper Saddle River, New Jersey: Prentice-Hall, Inc., 1993.

[11] P. Moulin and R. Koetter, "Data-hiding codes," *Proceedings of the IEEE*, vol. 93, no. 12, pp. 2083–2126, Dec. 2005.

[12] ETSI Rec. GSM 05.05, "Digital Cellular Telecommunication System (Phase 2+); Radio transmission and reception," version 8.9.0, 2001.

Acknowledgment — Special thanks to *Helge Lüders* for preparing and conducting the GSM link-level simulations.