Artificial Bandwidth Extension of Speech Supported by Watermark-Transmitted Side Information

Bernd Geiser, Peter Jax, and Peter Vary

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

Abstract

In recent years research in the area of *artificial bandwidth extension* (BWE) of speech signals has made significant progress. The intention of BWE is to produce wideband speech with a cut-off frequency of for instance 7 kHz from a narrowband version (e.g., with telephone bandwidth, i.e., $300\,\mathrm{Hz} - 3.4\,\mathrm{kHz}$). The respective algorithms are based on the estimation of parameters of a source model for speech production given the knowledge of the narrowband signal. A theoretical *performance bound* on this estimation has been formulated in [1]. In order to overcome this boundary we propose the transmission of (compact) *side information* which can support the parameter estimation. Since a common additional channel would conflict with the requirement of backwards compatibility in narrowband communication systems, the side information is embedded as *digital watermark* into the narrowband speech.

1. Introduction

The vast majority of today's telecommunication infrastructure still employs narrowband speech coding with a bandwidth of 3.1 kHz which has been chosen to guarantee a sentence intelligibility of about 99%. However, this typical "telephone sound" cannot satisfy increased demands. For instance, the intelligibility of single syllables can be improved by an extended speech bandwidth of 7 kHz. Thus, the listening effort is reduced, i.e., listening to the wideband speech is more comfortable.

Various wideband speech codecs have been developed in the past. A popular example is the Adaptive Multi-Rate Wideband (AMR-WB) codec. Though providing very good quality for the given bit rate, these codecs could not yet be established on large scale in fixed or mobile communication networks since their implementation requires changes both in terminals *and* within the network.

Techniques for *artificial bandwidth extension* (BWE) allow to leave the sending station and the communication network unchanged. Only the receiving station has to be modified. However, pure BWE suffers from inherently limited wideband speech quality as found by several authors [1], [2]. Giving up the backwards compatibility with respect to the sending station one can do better than this boundary by extracting suitable side information from the original speech signal and transmitting it over a separate channel. In order to retain backwards compatibility with respect to the communication network, this transmission channel can be realized by the embedding of a *digital watermark* into the narrowband speech. As an additional benefit, unmodified receivers are still usable as narrowband terminals.

2. Preliminaries

First, we will briefly review our stand-alone BWE algorithm. We will then define the *log spectral distortion* as a quality measure which is used for exemplary comparison of the performance of our BWE algorithm and of wideband speech coding.

2.1. Artificial Bandwidth Extension

The signal flow of the *stand-alone* BWE algorithm from [1] and [3] is depicted in Figure 1. After an *interpolation* of the narrowband signal to wideband sample rate, a *feature-vector* $\mathbf{x}_\mathrm{f}$ is computed. Then, by means of a statistical *hidden Markov model* (HMM), an estimate for the wideband spectral envelope is determined in terms of *linear prediction* (LP) coefficients $\mathbf{a}_{\mathrm{wb,BWE}}$. These wideband coefficients are used for *analysis*

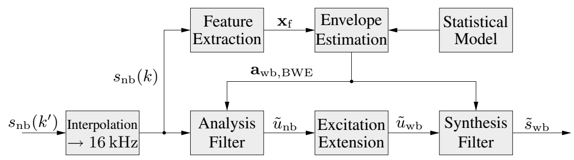

Figure 1: Stand-alone BWE algorithm.

filtering of the interpolated narrowband signal. After the *extension of the* resulting *excitation* (see [1], [3]), the inverse *synthesis filter* is applied. The choice of an excitation extension which does not alter the narrowband part leads to a BWE system which is *transparent* w.r.t. the *narrowband components*.

2.2. Instrumental Performance Measure

In this paper "speech quality" of bandwidth extended signals is assessed in terms of an objective *subband spectral distortion* measure, i.e., a distance measure for the *spectral envelopes* of the subband signals [1]:

$$d^2_{\mathrm{LSD,\,hb}} \doteq \frac{1}{2\pi}\int_{-\pi}^{\pi}\Big(20\lg\frac{\sigma_{\mathrm{rel}}\cdot|A_{\mathrm{hb,BWE}}(e^{j\Omega})|}{\sigma_{\mathrm{rel,BWE}}\cdot|A_{\mathrm{hb}}(e^{j\Omega})|}\Big)^2\,\mathrm{d}\Omega, \qquad (1)$$

where $\Omega \doteq \pi\cdot\mathrm{sgn}(f)\cdot(|f|-f_1)/(f_2-f_1)$, $f_1 \le |f| \le f_2$. Here the considered subband (hb $\hat{=}$ high band) is bounded by $f_1 = 3.4\,\mathrm{kHz}$ and $f_2 = 7\,\mathrm{kHz}$. A_{hb} represents the frequency response of the linear prediction filter for this band and σ_{rel} is the high band gain normalized by the narrowband gain. In practice, $d_{\mathrm{LSD,\,hb}}$ can be approximated by using the first 9 coefficients $c_0,\ldots,c_8$ of the real cepstrum in the high subband:

$$d^2_{\mathrm{LSD,hb}} \approx \left(\frac{10}{\ln 10}\right)^2\Big((c_0 - c_{0,\mathrm{BWE}})^2 + 2\sum_{i=1}^{8}(c_i - c_{i,\mathrm{BWE}})^2\Big). \quad (2)$$

Our stand-alone BWE algorithm, as outlined in Section 2.1, achieves mean spectral distortions down to about $6\,\mathrm{dB}$. On the other hand, the examination of the AMR-WB codec at $23.05\,\mathrm{kbit/s}$ — an example of "near transparent" wideband speech coding — gives a mean distortion of about $3\,\mathrm{dB}$. This value has been adopted here as our "target quality". The discrepancy of about $3\,\mathrm{dB}$ between this target quality and stand-alone BWE motivates the support of BWE with additional side information.

3. BWE with Side Information

The proposed system for BWE with side information [4] is illustrated in Figure 2. At the transmitting terminal, the high band spectral envelope of the wideband input signal is analyzed and the side information is determined (see Section 3.1). This side information (*message* m) is encoded by the watermark transmitter and added to the telephone bandpass ($300\,\mathrm{Hz} - 3.4\,\mathrm{kHz}$) filtered speech. Note that, using a well-designed watermarking scheme, the transmitted signal $\mathbf{s}^w_{\mathrm{nb}}$ should not be subjectively distinguishable from the original telephone speech $\mathbf{s}_{\mathrm{nb}}$. At the

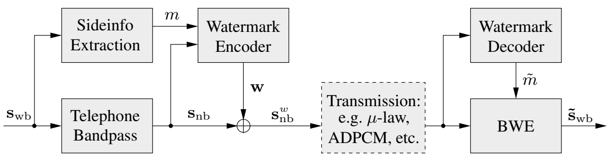

Figure 2: Overview of the transission system.

receiver, the watermark is extracted and the decoded side information message $\tilde{m}$ is used to control the BWE receiver block (see Section 3.2). The watermarking scheme will be described in Section 4.

3.1. Choice of Side Information

The feature extraction from Figure 1 lacks information on the upper band's spectral envelope. In order to further support the wideband envelope estimation we additionally extract a spectral representation of frequencies above $3.4\,\mathrm{kHz}$ from the *wideband* signal (which is only available at the sending side). This subband envelope is computed by *selective linear prediction*, i.e., computation of the wideband power spectrum followed by an IDFT of its upper band components and a subsequent Levinson-Durbin recursion of order 8. The resulting subband LPC coefficients are converted into the cepstral domain and are finally quantized by a vector quantizer with a codebook of size $M \doteq 2^N$. The training of this vector quantizer shows that a 6 bit codebook ($M = 64$) is sufficient to achieve a mean spectral distortion of less than $3\,\mathrm{dB}$ for the upper band which is our target "speech quality" (compare to Section 2.2).

The side information which is transmitted as watermark message m is the codebook index of the quantized cepstral vector for each speech frame:

$$m \in \mathbb{M} \doteq \{1, \ldots, M\}. \tag{3}$$

For a frame length of for instance $\tau = 20\,\mathrm{ms}$ this results in a *side information data rate* of $r_{\mathrm{s}} = 50 \cdot N\,\mathrm{bit/s}$.

3.2. Integration into the BWE Algorithm

There are several possibilities to make use of the side information as defined by (3) within the BWE algorithm. The most intuitive approach is to omit the envelope estimation in Figure 1 and to replace it with a block which assembles the narrowband envelope (available from the speech signal) and the upper band envelope (available as the side information m) and then outputs the desired coefficients $\mathbf{a}_{\mathrm{wb,BWE}}$.

Let $\mathbf{x}^w_{\mathrm{recv}}$ denote the received side information vector whose meaning and properties will be specified in Section 4.2. If a *likelihood function* $p(\mathbf{x}^w_{\mathrm{recv}}|m)$ is available at the receiver, the cepstral vector $\tilde{\mathbf{c}}$ which represents the upper band's envelope can, in analogy to [1] and [3], be estimated with either a *maximum likelihood*, a *maximum a posteriori* or a *minimum mean square error* rule. The **maximum likelihood** index m_{ML} is computed like this:

$$m_{\mathrm{ML}} = \arg\max_{m \in \mathbb{M}} p(\mathbf{x}^w_{\mathrm{recv}}|m). \tag{4}$$

Then $\tilde{\mathbf{c}}_{\mathrm{ML}} \doteq \hat{\mathbf{c}}_{m_{\mathrm{ML}}}$ is looked up in the vector quantization codebook. The **maximum a posteriori** (MAP) rule additionally takes a priori probabilities $P(m)$ into account:

$$m_{\mathrm{MAP}} = \arg\max_{m \in \mathbb{M}} \frac{P(m) \cdot p(\mathbf{x}^w_{\mathrm{recv}}|m)}{\sum\limits_{m' \in \mathbb{M}} P(m') \cdot p(\mathbf{x}^w_{\mathrm{recv}}|m')} \tag{5}$$

and the resulting vector is $\tilde{\mathbf{c}}_{\mathrm{MAP}} \doteq \hat{\mathbf{c}}_{m_{\mathrm{MAP}}}$. **MMSE** estimation essentially is a weighted sum of the M vector codebook centroids $\hat{\mathbf{c}}_m$:

$$\tilde{\mathbf{c}}_{\mathrm{MMSE}} \doteq \sum_{m \in \mathbb{M}} \hat{\mathbf{c}}_m \cdot \frac{P(m) \cdot p(\mathbf{x}^w_{\mathrm{recv}}|m)}{\sum\limits_{m' \in \mathbb{M}} P(m') \cdot p(\mathbf{x}^w_{\mathrm{recv}}|m')}. \tag{6}$$

Another approach is the *combination* of the side information with the feature extraction and the envelope estimation of the stand-alone BWE algorithm from Figure 1. In this case the likelihood function is substituted by the conditional joint PDF of the received side information $\mathbf{x}^w_{\mathrm{recv}}$ *and* the features $\mathbf{x}_{\mathrm{f}}$. For a **joint MMSE** estimation rule [4] this gives:

$$\tilde{\mathbf{c}}_{\text{J-MMSE}} \doteq \sum_{m \in \mathbb{M}} \hat{\mathbf{c}}_m \cdot \frac{P(m) \cdot p(\mathbf{x}^w_{\mathrm{recv}}, \mathbf{x}_{\mathrm{f}}|m)}{\sum\limits_{m' \in \mathbb{M}} P(m') \cdot p(\mathbf{x}^w_{\mathrm{recv}}, \mathbf{x}_{\mathrm{f}}|m')}. \tag{7}$$

The conditional joint PDF can be assumed to be separable:

$$p(\mathbf{x}^w_{\mathrm{recv}}, \mathbf{x}_{\mathrm{f}}|m) = p(\mathbf{x}^w_{\mathrm{recv}}|m) \cdot p(\mathbf{x}_{\mathrm{f}}|m), \tag{8}$$

where the first factor is our likelihood function and the latter one is an integral part of the envelope estimation block and thus available within the BWE algorithm.

4. Watermarking Scheme

We will now address the actual watermark embedding of the side information — as defined in Section 3.1 — into the narrowband speech signal s_{nb}. Also the message extraction at the receiver side will be accounted for.

4.1. Watermarking — Basics

The process of watermark embedding into a *host signal* s can be described as addition of a so called *watermark signal* w which is dependent on the *message* m to be transmitted. In order to eliminate influences of the host signal on the watermark decoding process, additional dependencies on the signal s itself can be introduced. So a generic embedding function reads:

$$f_{\mathrm{wm}}(s, m) = s + w(s, m). \tag{9}$$

Two power ratios are important in the context of watermark embedding. The **Host to Watermark Ratio** (HWR) is a measure for the embedding power of the watermark:

$$\mathrm{HWR} \doteq \sigma_s^2/\sigma_w^2 \quad \text{or} \quad \mathrm{HWR_{dB}} \doteq 10\lg\sigma_s^2/\sigma_w^2\,\mathrm{dB}. \quad (10)$$

It should be kept as high as possible to guarantee a negligible perceptibility of the watermark w. The **Watermark To Noise Ratio** (WNR) characterizes the *watermark channel* quality[1]:

$$\mathrm{WNR} \doteq \sigma_w^2/\sigma_n^2 \quad \text{or} \quad \mathrm{WNR_{dB}} \doteq 10\lg\sigma_w^2/\sigma_n^2\,\mathrm{dB}. \quad (11)$$

Error rates and parameter SNRs will be plotted over the WNR.

4.2. Embedding the Watermark

The watermark is embedded into *signal vectors* $\mathbf{x}$ which are *not* taken from the speech signal itself but from a transformed domain. Section 4.4 introduces the respective transformation $\mathcal{T}:\mathbf{s}_{\mathrm{nb}}\to\mathbf{x}$ and its inverse $\mathcal{T}^{-1}:\mathbf{x}\to\mathbf{s}_{\mathrm{nb}}$.

Our choice for the (now vectorial) embedding function $\mathbf{f}_{\mathrm{wm}}(\mathbf{x},m)$ is a specialization of Chen's *Quantization Index Modulation* (QIM) as described in [5]. In the QIM embedding process the message $m\in\mathbb{M}$ selects one of M vector-quantizers which is then applied to the n-dimensional vector $\mathbf{x}$:

$$\mathbf{f}_{\mathrm{wm,QIM}}(\mathbf{x},m) = \mathcal{Q}_m(\mathbf{x}) \quad (12)$$

resulting in a watermark data rate of $r_{\mathrm{wm}} = \mathrm{ld}\,M\ ^{\mathrm{bit}}/_{\mathrm{vector}}$. The quantizers $\mathcal{Q}_m$ are designed such that the index m can be inferred from the knowledge of the quantized vector, i.e., the M sets of centroids are non-intersecting:

$$\mathcal{Q}_i(\mathbf{x}) \neq \mathcal{Q}_j(\mathbf{x}) \quad \forall\mathbf{x},\ i,j\in\mathbb{M} \text{ and } i\neq j. \quad (13)$$

Here, the M vector-quantizers are chosen as shifted versions of a prototype *lattice-quantizer* $\mathcal{Q}_{\Lambda_\mathrm{c}}$ associated to a lattice Λ_c with good quantization properties[2]. Hence, we call our watermarking scheme Λ-QIM. The choice of shifted versions of a prototype quantizer ensures that the watermark power is independent of the message m and only given by the quantization performance of $\mathcal{Q}_{\Lambda_\mathrm{c}}$[3]. Further, it gives rise to an efficient implementation:

$$\begin{aligned}\mathbf{f}_{\mathrm{wm},\Lambda\text{-QIM}}(\mathbf{x},m) &= \mathcal{Q}_m(\mathbf{x}) \\ &= \mathcal{Q}_{\Lambda_\mathrm{c}}(\mathbf{x}-\mathbf{d}_m)+\mathbf{d}_m. \quad (14)\end{aligned}$$

The vectors $\mathbf{d}_m$ (w.l.o.g. $\mathbf{d}_1=\mathbf{0}$) are called *shift vectors* (or in lattice theory *coset leaders*) and have been found by sublattice-decomposition of a *fine lattice* Λ_f or — as an alternative — with a LBG-style training algorithm[4]. The optimization criterion for the shift vectors $\mathbf{d}_m$ is the minimization of the decoding error probability for given channel distortions.

Additional improvements for (14) are, e.g., described in [5], [6] and [7] and include so called *precoding* (in [5] *distortion compensation*) and *dithering*. Note that Λ-QIM is similar to the watermarking scheme from [6] with a different construction of the shift vector codebook. Our choice leads to so called *nested lattice codes* as introduced, e.g., in [7].

[1] Note that in this context the term *noise* does not only refer to channel noise but to all distortions such as those introduced by the speech codec.

[2] Λ_c stands for *coarse lattice*.

[3] Under the assumption that $(\mathbf{x} \bmod \Lambda_\mathrm{c}) \doteq \mathbf{x}-\mathcal{Q}_{\Lambda_\mathrm{c}}(\mathbf{x})$ is uniformly distributed over the basic Voronoi cell of Λ_c.

[4] The latter does not ensure that the union of all shifted lattices is a lattice itself but rather generates a generic *tiling*.

4.3. Decoding the Watermark

The input of the watermark decoder is the watermarked signal vector $\mathbf{x}^w$ plus an effective noise $\mathbf{n}_{\mathrm{eff}}$ and is denoted $\mathbf{x}^w_{\mathrm{recv}}$:

$$\begin{aligned}\mathbf{x}^w_{\mathrm{recv}} &= \mathbf{x}^w+\mathbf{n}_{\mathrm{eff}} \\ &= \mathbf{f}_{\mathrm{wm},\Lambda\text{-QIM}}(\mathbf{x},m)+\mathbf{n}_{\mathrm{eff}}. \quad (15)\end{aligned}$$

First, the quantization error is computed for each of the M quantizers which is then used to produce an estimate for the embedded message m. For all $m\in\mathbb{M}$ the squared quantization error can be formally written as a euclidian metric which the decoding process is based on:

$$\begin{aligned}\Delta_m &= \left|\mathbf{x}^w_{\mathrm{recv}} \bmod (\Lambda_\mathrm{c}+\mathbf{d}_m)\right|^2 \\ &= \left|\mathbf{x}^w_{\mathrm{recv}} - \mathcal{Q}_{\Lambda_\mathrm{c}+\mathbf{d}_m}(\mathbf{x}^w_{\mathrm{recv}})\right|^2 \\ &= \left|\mathbf{x}^w_{\mathrm{recv}} - [\mathcal{Q}_{\Lambda_\mathrm{c}}(\mathbf{x}^w_{\mathrm{recv}}-\mathbf{d}_m)+\mathbf{d}_m]\right|^2. \quad (16)\end{aligned}$$

A simple *hard decision* estimator would select the quantizer giving the minimum metric which (for non-degenerated channels) is equivalent to the *maximum likelihood* estimation rule from equation (4):

$$m_{\mathrm{ML}} = \arg\min_{m\in\mathbb{M}}\Delta_m. \quad (17)$$

More sophisticated rules like MAP (5) or (joint) MMSE (6),(7) demand a *likelihood function* $p(\mathbf{x}^w_{\mathrm{recv}}|m) = p(\mathbf{x}^w_{\mathrm{recv}}|\mathbf{d}_m)$ with $m\in\mathbb{M}$ which takes channel characteristics into account, e.g., for a (equivalent) i.i.d. gaussian channel:

$$p(\mathbf{x}^w_{\mathrm{recv}}|\mathbf{d}_m) \approx \frac{1}{\sqrt{(2\pi\tilde{\sigma}^2)^n}}\exp\left(-\frac{\Delta_m}{2\tilde{\sigma}^2}\right), \quad (18)$$

where $\tilde{\sigma}$ is an estimate for the standard deviation of the effective noise $\mathbf{n}_{\mathrm{eff}}$ and n is the dimension of the signal vectors.

4.4. Subspace Embedding and Masking

Now the transformation $\mathcal{T}:\mathbf{s}_{\mathrm{nb}}\to\mathbf{x}$ and its inverse will be described. The actual watermarking process at the *encoder side* is done in a scaled subspace of the LPC filtered speech (i.e., the *excitation signal* $\mathbf{e}$). Afterwards, the inverse operations are performed in reversed order: inverse scaling, backprojection and synthesis filtering. An overview of this encoder is presented in Figure 3. At the *decoder side* of the considered system another identical transformation $\mathcal{T}$ has to be implemented in order to get access to the Λ-QIM watermarked signal $\mathbf{x}^w$.

The effect of the *LPC analysis/synthesis* filtering can be interpreted as *noise shaping* of the watermark signal which then follows the spectral characteristic of the speech (also compare to [8]). *Subspace projection* of the LPC-filtered signal $\mathbf{e}$ is applied in order to reduce the signal vector dimension from the speech frame size (160 samples) to the dimension of the lattice Λ_c which is used for quantization in the Λ-QIM[5]. A careful selection of the orthonormal subspace basis or, respectively, of the projection matrix $\mathbf{\Phi}\in\mathbb{R}^{n_\mathbf{e}\times n}$ (with $n_\mathbf{e}=160$, $n_\mathbf{e}>n$) provides an additional way of influencing the perceptibility *and* the robustness of the watermark. Finally, the value of the *scale factor* c adjusts the watermark "embedding strength", i.e., the HWR, thus avoiding any scaling of the lattice Λ_c and of the shift vectors $\mathbf{d}_m$.

[5] It is also possible to do a "multi-subspace" embedding which, due to lower dimensional lattices, contributes to complexity reduction.

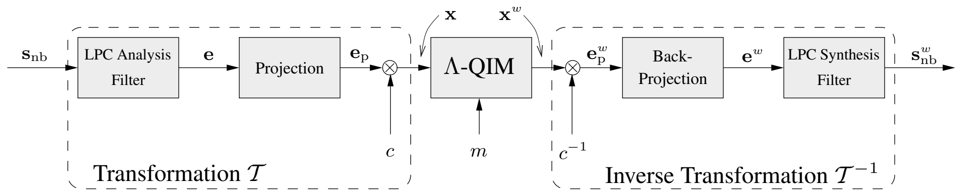

Figure 3: Overview of the watermark embedding in connection with the transformation $\mathcal{T}$ and its inverse $\mathcal{T}^{-1}$.

5. Simulation Results

Our simulation scenario includes the Λ-QIM from Sections 4.2 and 4.3 for different lattices and code constructions. Here we present results for $\Lambda_c = \Lambda_{16}$, i.e., the 16-dimensional *Barnes-Wall-Lattice*, and, according to Section 3.1, for a data rate of $300\,{}^{\mathrm{bit}}/_{\mathrm{s}} = 6\,{}^{\mathrm{bit}}/_{\mathrm{frame}}$. So the codebook size is $M = 2^6 = 64$. The Λ-QIM shift vectors have been found with LBG training[6] and the actual quantization has been performed with the algorithm from [9]. A transformation including projection and LPC filtering as described in Section 4.4 has been used.

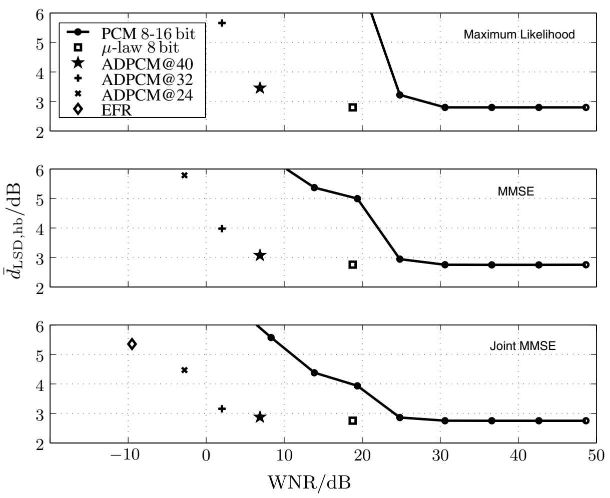

Figure 4: Mean subband spectral distortion $\bar{d}_{\mathrm{LSD,hb}}$ for the estimation rules from equations (4), (6) and (7) and different transmission channels. In the Λ-QIM the Barnes-Wall-Lattice Λ_{16} has been used as Λ_c.

The watermark embedding has been done with a HWR of about 20 dB which — in conjunction with the transformation $\mathcal{T}$ (Section 4.4) — leads to a nearly imperceptible watermark. The watermarked signal has been exposed to various channel distortions (transcoding):

- PCM — scalar uniform quantization (8 to 16 bit),
- μ-law — scalar nonuniform quantization (8 bit),
- ADPCM — ITU-T G.726 (24, 32, and 40 kbit/s),
- EFR — ETSI-GSM 06.60 (12.2 kbit/s).

In Figure 4 these results are presented in terms of the subband spectral distortion measure for different estimation rules in the bandwidth extension algorithm. $\bar{d}_{\mathrm{LSD,hb}}$ has been calculated using (2) and is plotted over the WNR (11).

[6]The nested lattice codes derived by our sublattice decomposition yield similar results.

6. Conclusion

We have introduced a system for artificial bandwidth extension of speech signals which benefits from the transmission of additional *side information* such that better performance is achieved than the theoretical bound for pure feature based BWE suggests. The side information is transmitted in a backwards compatible manner, i.e., as a *digital watermark*.

With our system the usual narrowband quality is retained whereas a specially equipped receiving terminal with watermark decoder and a bandwidth extension algorithm can produce wideband speech of considerable quality. Measurements of the mean subband spectral distortion $\bar{d}_{\mathrm{LSD,hb}}$ for different transcodings (Figure 4) show a clear performance gain over stand-alone BWE (which yields values of $\bar{d}_{\mathrm{LSD,hb}} \approx 6\,\mathrm{dB}$). In many relevant cases, for example ISDN (μ-law) and DECT (ADPCM), the distortion measure is now comparable to wideband speech codecs like the AMR-WB, i.e., $\bar{d}_{\mathrm{LSD,hb}} \approx 3\,\mathrm{dB}$.

7. References

[1] P. Jax, “Enhancement of bandlimited speech signals: Algorithms and theoretical bounds,” Ph.D. dissertation, RWTH Aachen University, Aachen, Germany, 2002.

[2] M. Nilsson, H. Gustafsson, S. V. Andersen, and W. B. Kleijn, “Gaussian mixture model based mutual information estimation between frequency bands in speech,” in *Proc. of ICASSP*, vol. 1, Orlando, FL, USA, May 2002, pp. 525–528.

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

[4] B. Geiser, “Backwards compatible wideband speech communication with artificial bandwidth extension and watermarking,” Diplomarbeit, RWTH Aachen, Institut für Nachrichtengeräte und Datenverarbeitung, 2004, (in German).

[5] B. Chen and G. W. Wornell, “Quantization index modulation: A class of provably good methods for digital watermarking and information embedding,” *IEEE Trans. Inform. Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.

[6] R. F. H. Fischer, R. Tzschoppe, and R. Bäuml, “Lattice Costa scheme using subspace projection for digital watermarking,” in *Proc. of Intl. ITG Conf. on Source and Channel Coding*, Erlangen, Germany, Jan. 2004, pp. 127–134.

[7] R. Zamir, S. Shamai, and U. Erez, “Nested linear/lattice codes for structured multiterminal binning,” *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1250–1276, June 2002.

[8] Q. Cheng and J. Sorensen, “Spread spectrum signaling for speech watermarking,” in *Proc. of ICASSP*, vol. 3, Salt Lake City, UT, USA, May 2001, pp. 1337–1340.

[9] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, “Closest point search in lattices,” *IEEE Trans. Inform. Theory*, vol. 48, no. 8, pp. 2201–2214, Aug. 2002.