AN UPPER BOUND ON THE QUALITY OF ARTIFICIAL BANDWIDTH EXTENSION OF NARROWBAND SPEECH SIGNALS

Peter Jax and Peter Vary

Institute of Communication Systems and Data Processing,
Aachen University of Technology, Templergraben 55, D-52056 Aachen, Germany
E-mail: jax@ind.rwth-aachen.de

ABSTRACT

The aim of the artificial bandwidth extension (BWE) of speech signals is to recover wideband speech from bandlimited speech. As the BWE algorithm is supposed to operate without additional side information on the original wideband speech, it has to exploit mutual dependencies between the available and missing frequency bands of the speech signal.

In this paper the BWE is examined from an information theoretic perspective. After defining a performance measure, and introducing a few assumptions on a generalized BWE algorithm, a general relationship between mutual information and the maximum achievable estimation performance is formulated, which ensues an upper bound on the performance of BWE algorithms. Finally, some measurements considering a representative BWE scenario are presented.

1. INTRODUCTION

In current public telephone systems the bandwidth of the transmitted speech is limited due to constraints of the old analogue telephone system to a frequency range of up to about 3.4 kHz. This bandwidth limitation causes the characteristic sound of "telephone speech". Listening experiments have shown that an increased frequency bandwidth of speech signals contributes significantly to the perceived speech quality as well as to intelligibility.

True wideband speech communication requires a modification of the transmission link by enhanced speech codecs. An alternative approach towards a higher (audio) bandwidth is the artificial bandwidth extension (BWE): missing low and high frequency components of the speech signal are recovered at the receiving end of the transmission link utilizing only the bandlimited speech.

It is plausible that an extension of the bandwidth of speech signals is only possible, if there are sufficient dependencies between the available bandlimited speech signal and the missing frequency components. The fact that the narrowband speech and the missing signal components are results of the same physical speech production process gives rise to the assumption that there are such dependencies in speech signals. This assumption is supported by the success of many methods for BWE published throughout the last decade (e.g. [1] – [7]). There are very few publications so far which shade some light onto the information theoretic background of artificial bandwidth extension [8], [9].

In digital signal processing, the *linear* dependencies between signals are commonly described in terms of *correlation* factors. In an information theoretic perspective, the dependencies between different signals are described by their mutual information (MI). In contrast to the correlation measure, the MI covers all kinds of linear and non-linear dependencies. The aim of this paper is to investigate the relationship between the maximum achievable quality of a BWE algorithm on one hand and the mutual information between representations of the bandlimited speech and of the missing frequency components on the other hand.

2. ARTIFICIAL BANDWIDTH EXTENSION

Within the scope of this paper, we will concentrate on the extension of the bandwidth of low-pass filtered speech signals towards high frequencies. According to the frequency range of typical telephone speech, we assume that the narrowband speech signal contains frequency components up to a cutoff frequency of about 3.4 kHz — this frequency band will be called the *base-band* in the following. The aim of the bandwidth extension algorithm is the recovery of signal components up to a cutoff frequency of 7 kHz, which is the upper band limit of typical *wideband* speech. The extended frequency range between 3.5 and 7 kHz will be called *missing* or *extended* frequency band in the following. Note that the results of this paper can be applied to different BWE scenarios as well, e.g., to the bandwidth extension toward low frequencies.

The vast majority of the adaptive BWE algorithms published in literature to date are based on the well-known linear source-filter model of the speech production process: it is assumed that the human vocal tract can be modeled by an auto-regressive (AR) filter ($1/A(z)$), which is excited by a spectrally flat excitation signal ($u(k)$). Consequently, the bandwidth extension of the speech signal is commonly performed separately for the spectral envelope and the excitation of the speech [1], [2] (cp. Fig. 1). It has been found that the spectral envelope is particularly important for the subjective quality of the extended speech (e.g. [2]).

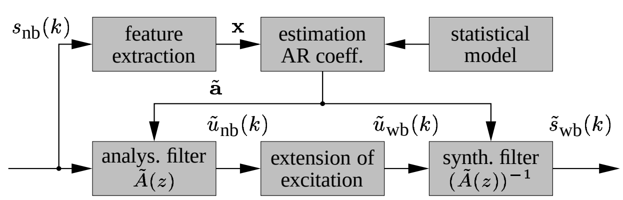

Fig. 1. Signal-flow of an exemplary BWE algorithm (from [6]).

2.1. Performance Measure

For evaluating the performance of a BWE algorithm, the distortion of the spectral envelope of the extended speech signal shall be measured w.r.t. the original wideband speech signal. Since the base-band speech is readily available in the BWE system, we are interested in the *shape* as well as the *gain* of the spectral envelope in the missing frequency range only. The gain shall be expressed with respect to the base-band signal components, i.e., it shall be a *relative* gain as specified later (cp. e.g. [8], [7]).

To define the spectral distortion measure, first the wideband speech signal is split into two sub-band signals, which contain only the base-band components, or only the extended frequency components of the wideband speech respectively. For this purpose the wideband speech is resampled to a sampling rate of 14 kHz. Then the two sub-band signals are determined via low-pass filtering (cutoff frequency 3.4 kHz) and high-pass filtering (cutoff frequency 3.5 kHz) respectively. Finally, the two signals are downsampled by a factor of two by omitting every second sample of the signals. The two resulting sub-band signals are denoted by $s_{\mathrm{bb}}(k)$ for the base-band signal, and by $s_{\mathrm{mb}}(k)$ for the signal containing the missing frequency band of the wideband speech. Both sub-band signals are sampled at a sampling rate of 7 kHz.

In the next step, two individual auto-regressive models are fitted to frames (duration 20 ms) of the two sub-band signals. This is accomplished by a conventional LPC analysis, i.e., by calculation of the auto-correlation coefficients and a subsequent *Levinson-Durbin* algorithm [10]. The model used for representing the missing frequency band spectrum $|S_{\mathrm{mb}}(e^{j\omega})|^2$ is, e.g., given by $\sigma_{\mathrm{mb}}^2/|A_{\mathrm{mb}}(e^{j\omega})|^2$ (cp. Fig. 2). The results of the LPC analysis are the coefficient set $\mathbf{a}_{\mathrm{mb}}$ (of $A_{\mathrm{mb}}(z)$) representing the spectral envelope of the missing frequency band, as well as two gain factors σ_{bb} and σ_{mb} of the base-band and the missing frequency band respectively. Since the relative gain of the extended frequency band shall be measured, we define $\sigma_{\mathrm{rel}} = \sigma_{\mathrm{mb}}/\sigma_{\mathrm{bb}}$. The order of $\mathbf{a}_{\mathrm{mb}}$ is N_a.

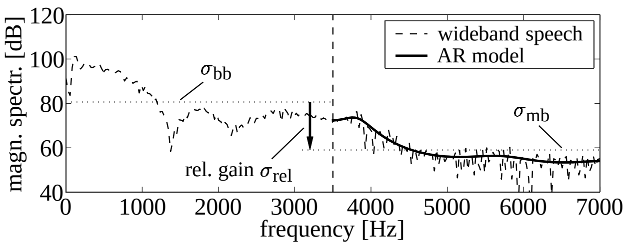

Fig. 2. AR modeling of the spectral envelope of the sub-band signal containing the missing frequency band.

We define the performance of the estimation of the wideband spectral envelope in terms of the *log spectral distortion* (LSD) in the missing frequency band

$$d_{\mathrm{LSD}}^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(20\log_{10}\frac{\sigma_{\mathrm{rel}}}{|A_{\mathrm{mb}}(e^{j\omega})|} - 20\log_{10}\frac{\tilde{\sigma}_{\mathrm{rel}}}{|\tilde{A}_{\mathrm{mb}}(e^{j\omega})|}\right)^2 d\omega, \quad (1)$$

where the quantities $A_{\mathrm{mb}}(e^{j\omega})$ and σ_{rel} label the modeled frequency spectrum and relative gain of the missing frequency band of the original wideband speech, and $\tilde{A}_{\mathrm{mb}}(e^{j\omega})$ and $\tilde{\sigma}_{\mathrm{rel}}$ denote the corresponding estimated parameters as determined by a bandwidth extension system[1]. Note that the LSD measure is evaluated for the sub-band signal $s_{\mathrm{mb}}(k)$ containing the missing frequency band only, such that the integration range of $-\pi$ to π in (1) covers the missing frequency range of 3.5 to 7 kHz in the original wideband speech signal. The unit of d_{LSD} is dB.

Using *Parseval*'s theorem, the LSD can alternatively be expressed in cepstral domain [11]

$$d_{\mathrm{LSD}}^2 = \left(\frac{10}{\log_e 10}\right)^2 \left((c_0-\tilde{c}_0)^2 + 2\sum_{i=1}^{\infty}(c_i-\tilde{c}_i)^2\right), \quad (2)$$

where the cepstral coefficients $c_0, c_1, \ldots$ are calculated from the AR coefficients $\mathbf{a}_{\mathrm{mb}}$ and the relative gain σ_{rel} of the missing frequency band via a recursive formula given in [10]. Only the first N_a+1 cepstral coefficients derived in this way are non-redundant, while the remaining coefficients can be determined from these foremost coefficients. Note that the 0-th coefficient c_0 is not set to zero here, as it is often the case if only the shape of the spectrum shall be evaluated. For a sequence of speech frames, the *root mean-square* (RMS) average of the LSD is given by

$$\bar{d}_{\mathrm{LSD}} = \frac{\sqrt{2}\,10}{\log_e 10}\sqrt{E\left\{\frac{1}{2}(c_0-\tilde{c}_0)^2 + \sum_{i=1}^{\infty}(c_i-\tilde{c}_i)^2\right\}}. \quad (3)$$

The LSD measure correlates reasonably well with the subjective speech quality. Therefore, it has found wide acceptance, for example in the area of speech coding, where it is commonly used to assess the quality of quantizers of the spectral envelope. As a rule of thumb, the speech quality is called sufficient (or "transparent" in the case of (narrowband) speech coding), if the log spectral distortion is less than 1 dB on the average. Usually, the number of outliers with a high distortion (e.g. > 2 dB) is also evaluated.

2.2. Assumptions on the BWE Algorithm

The estimation of the shape of the spectral envelope in a BWE system will be described in the following by the simple function

$$\tilde{\mathbf{y}} = f(\mathbf{x}), \quad \text{with the estimation error} \quad \mathbf{n} = \mathbf{y} - \tilde{\mathbf{y}}. \quad (4)$$

Because the performance of the BWE algorithm shall be measured in terms of the LSD measure from the previous section, the quantity $\mathbf{y}$ resembles a *weighted* cepstral representation of the missing frequency band, which can be determined from cepstral coefficients $c_0, c_1, \ldots c_{d-1}$ as follows

$$y_i = \begin{cases} \frac{1}{\sqrt{2}}c_i, & \text{if } i = 0 \\ c_i, & \text{if } 1 \le i < d. \end{cases} \quad (5)$$

The scalar values y_i constitute the d-dimensional vector $\mathbf{y} = [y_0, y_1, \ldots y_{d-1}]^T$. Note that the quantity $\tilde{\mathbf{y}}$ can be calculated for *any* BWE algorithm either from the extended speech signal or directly from the estimated representation of the wideband spectral envelope (as e.g. $\tilde{\mathbf{a}}$ in Fig. 1) if applicable.

For each frame of the input speech signal, the BWE system determines an individual estimate $\tilde{\mathbf{y}}$ on the basis of characteristics of the bandlimited input speech. These characteristics are represented by a feature vector $\mathbf{x}$, which is extracted from each frame of the input speech. In most BWE systems described in literature to date, the estimate of the spectral envelope is based on information on the spectral envelope of the narrowband speech signal, e.g., represented by its power spectral density or cepstral coefficients.

We have to impose the limitation here that the estimator is *memory-less*, i.e., the estimate $\tilde{\mathbf{y}}$ is calculated deterministically from the feature vector $\mathbf{x}$ for each individual signal frame. Note, however, that we do not assume anything about the particular realization of the estimator: the function $f(\mathbf{x})$ can be interpreted as a generalized description of any kind of linear or nonlinear estimator with arbitrary complexity — including the approaches commonly used for BWE, e.g., codebook / linear mapping, statistical estimation etc.

3. AN UPPER BOUND ON THE ESTIMATION PERFORMANCE

In this section a relation is derived between the mutual information $I(\mathbf{x};\mathbf{y})$ (between the features $\mathbf{x}$ of the narrowband speech signal and the representation $\mathbf{y}$ of the missing frequency band) and the maximum possible quality of an estimation $\tilde{\mathbf{y}}$ of the coefficients $\mathbf{y}$. The resulting bound is a generalization of a bound on the prediction gain for stationary scalar stochastic processes as described by *Bernhard* in [12]. *Bernhard*'s bound is reformulated here for the general scalar estimation problem, and it is extended to describe the estimation of *vector* signals.

[1] Estimated quantities are marked with a tilde (as in $\tilde{\sigma}_{\mathrm{rel}}$) in this paper.

3.1. Scalar Estimation

Let's first consider the case that the signal y to be estimated is a continuous scalar quantity (cp. [12]). The mutual information between y and $\mathbf{x}$ can be expressed with respect to the non-conditional and conditional differential entropies of y [13]

$$I(\mathbf{x}; y) = h(y) - h(y|\mathbf{x}) = h(y) - h(f(\mathbf{x}) + n|\mathbf{x}). \quad (6)$$

Because the estimate $f(\mathbf{x})$ is deterministically depending on $\mathbf{x}$, the conditional entropy $h(f(\mathbf{x}) + n|\mathbf{x})$ is equal to the conditional entropy $h(n|\mathbf{x})$ of the error signal due to the shift invariance of the entropy [12], [13]. In other words, the estimate $f(\mathbf{x})$ does not yield any new information, if the vector $\mathbf{x}$ is already known. Further, it is well known that conditioning reduces entropy, i.e., we have $h(n|\mathbf{x}) \leq h(n)$. Finally, the entropy of a scalar signal with the variance σ_n^2 is upper bounded by the entropy of a normally distributed variable with the same variance [13]. Applying the three statements of this paragraph in the same order yields the inequality

$$h(f(\mathbf{x}) + n|\mathbf{x}) = h(n|\mathbf{x}) \leq h(n) \leq \frac{1}{2} \log \left(2\pi e\, \sigma_n^2\right). \quad (7)$$

Note that logarithms are always to the base of e throughout this paper unless explicitly stated differently. Then, the unit of the differential entropies and mutual information is in nats. Using (7), Eq. (6) can be transformed to

$$I(\mathbf{x}; y) \geq h(y) - \frac{1}{2} \log \left(2\pi e\, \sigma_n^2\right). \quad (8)$$

3.2. Vector Estimation

The result from Sec. 3.1 shall now be generalized to the estimation of vectors. Hence, the variables $\mathbf{y}$ and $\mathbf{n}$ are now d-dimensional vectors. The estimation can be described individually for each component y_i of the vector $\mathbf{y}$, such that $\tilde{y}_i = f_i(\mathbf{x})$ and $n_i = y_i - \tilde{y}_i$. The estimation errors of the individual estimators can be calculated from (8)

$$\log \sigma_{n_i}^2 \geq 2 \left(h(y_i) - I(\mathbf{x}; y_i)\right) - \log (2\pi e) \quad (9)$$

or, if the exp operation is applied to both sides of the inequality

$$\sigma_{n_i}^2 \geq \frac{1}{2\pi e} \exp(2\, h(y_i|\mathbf{x})). \quad (10)$$

Adding up the inequalities of the individual estimators of all vector components yields a bound on the overall mean-square error

$$E\{|\mathbf{n}|^2\} \geq \sum_{i=0}^{d-1} \sigma_{n_i}^2 \geq \frac{1}{2\pi e} \sum_{i=0}^{d-1} \exp(2\, h(y_i|\mathbf{x})). \quad (11)$$

The first equality in (11) holds, if the estimate $\tilde{\mathbf{y}}$ is without bias, i.e., if $E\{f(\mathbf{x})\} = E\{\mathbf{y}\}$.

In the next step, we use a relation between the arithmetic and geometric mean of a sequence: if the numbers of a sequence $a_0, a_1, \ldots a_{N-1}$ are positive, the arithmetic mean is always greater than or equal to the geometric mean, i.e., $\frac{1}{N}(a_0 + a_1 + \ldots + a_{N-1}) \geq \sqrt[N]{a_0 a_1 \ldots a_{N-1}}$. In our case, the values of $\exp(2\, h(y_i|\mathbf{x}))$ are strictly positive and we find

$$\begin{aligned} E\{|\mathbf{n}|^2\} &\geq \frac{d}{2\pi e} \left(\prod_{i=0}^{d-1} \exp(h(y_i|\mathbf{x}))\right)^{\frac{2}{d}} \\ &= \frac{d}{2\pi e} \exp\left(\frac{2}{d} \sum_{i=0}^{d-1} h(y_i|\mathbf{x})\right). \end{aligned} \quad (12)$$

This expression contains the sum of the conditional entropies of the elements of $\mathbf{y}$. As it is well known, the joint entropy of several variables is always smaller than or equal to the sum of the individual entropies, i.e., $\sum_{i=0}^{d-1} h(y_i|\mathbf{x}) \geq h(y_0, y_1, \ldots y_{d-1}|\mathbf{x}) = h(\mathbf{y}|\mathbf{x})$, and therefore

$$\begin{aligned} E\{|\mathbf{n}|^2\} &\geq \frac{d}{2\pi e} \exp\left(\frac{2}{d}\, h(\mathbf{y}|\mathbf{x})\right) \\ &= \frac{d}{2\pi e} \exp\left(\frac{2}{d}\big(h(\mathbf{y}) - I(\mathbf{x}; \mathbf{y})\big)\right). \end{aligned} \quad (13)$$

Hence, a lower bound on the mean-square estimation error can be formulated in dependence of the mutual information $I(\mathbf{x}; \mathbf{y})$.

3.3. Application to Bandwidth Extension

The derived bound on the estimation performance shall now be applied to the problem of artificial bandwidth extension. First the mean-square error of the assumed spectral envelope estimator from Sec. 2.2 is investigated. Due to the weighting of the representation $\mathbf{y}$ of the missing frequency band, the MSE $E\{|\mathbf{n}|^2\}$ resembles a truncated version of the cepstral distance within the square root of (3). Because the truncated elements are non-negative, we find the inequality

$$\bar{d}_{\mathrm{LSD}} \geq \frac{\sqrt{2}\, 10}{\log 10} \sqrt{E\{|\mathbf{n}|^2\}}. \quad (14)$$

Hence, in a first step we can formulate a lower bound on the mean log spectral distortion of the estimated spectral envelope of the missing frequency band in dependence of the MSE of $\tilde{\mathbf{y}}$. The MSE of the estimator is further lower bounded by (13). Accordingly, inserting (13) into (14) yields

$$\bar{d}_{\mathrm{LSD}} \geq \underbrace{\frac{\sqrt{2}\, 10}{\log 10} \sqrt{\frac{d}{2\pi e}}}_{\triangleq\, C_d} \exp\left(\frac{1}{d}\big(h(\mathbf{y}) - I(\mathbf{x}; \mathbf{y})\big)\right), \quad (15)$$

which gives a lower bound on the achievable RMS log spectral distortion in dependence of the mutual information $I(\mathbf{x}; \mathbf{y})$ and of the differential entropy $h(\mathbf{y})$. Alternatively, a lower bound on the mutual information can be formulated, if a particular value of $\bar{d}_{\mathrm{LSD}}$ is specified

$$I(\mathbf{x}; \mathbf{y}) \geq h(\mathbf{y}) - d \log \frac{\bar{d}_{\mathrm{LSD}}}{C_d}. \quad (16)$$

Thus, if the actual MI $I(\mathbf{x}; \mathbf{y})$ is lower than the value on the right hand side of (16), it is by all means impossible to achieve the desired RMS log spectral distortion $\bar{d}_{\mathrm{LSD}}$.

4. MEASUREMENTS

In this section, the application of the bandwidth extension to recover frequencies above 3.5 kHz is considered. The missing upper frequency band between 3.5 and 7 kHz is represented by $d = 9$ (weighted) cepstral coefficients which are derived from $N_a = 8$ LPC coefficients $\mathbf{a}_{\mathrm{mb}}$ and the relative gain σ_{rel} as described in Sec. 2.1. For this particular application, the lower bound on the achievable mean log spectral distortion shall be estimated by approximating the differential entropy $h(\mathbf{y})$.

To define an estimation method for the differential entropy $h(\mathbf{y})$, we start with its integral definition

$$h(\mathbf{y}) = -\int \cdots \int p(\mathbf{y}) \log p(\mathbf{y})\, d\mathbf{y} = -E\{\log p(\mathbf{y})\}. \quad (17)$$

The d-dimensional integral in (17) can be interpreted as an expectation operation. This expectation operation can be approximated, if a model of the probability density function (PDF) $p(\mathbf{y})$ is available. In our investigations, we employed a *Gaussian mixture model* (GMM) for this purpose, which was constituted from $L = 256$ weighted multivariate *Gaussian* densities $\mathcal{N}(\cdot)$ with mean vectors μ_l and *full* covariance matrices Σ_l

$$p_m(\mathbf{y}) = \sum_{l=0}^{L-1} \rho_l \, \mathcal{N}(\mathbf{y}; \mu_l, \Sigma_l) \approx p(\mathbf{y}). \tag{18}$$

The scalar weights ρ_l as well as the parameters μ_l and Σ_l of the individual *Gaussians* were trained by the common *expectation-maximization* (EM) algorithm [14], [15]. The training data set consisted of more than 500000 non-overlapping frames of length 20 ms of wideband speech spoken by a number of different speakers in different languages.

With help of the model of the PDF $p(\mathbf{y})$, the result of the expectation operation from (17) can be approximated by numerical integration [15]

$$h(\mathbf{y}) \approx -\frac{1}{N_\nu} \sum_{\nu=0}^{N_\nu - 1} \log p_m(\mathbf{y}_m(\nu)). \tag{19}$$

The evaluation of the expectation operation is based on "synthetic" data vectors $\mathbf{y}_m(\nu)$, $\nu = 0, 1, \ldots N_\nu - 1$, which are generated by a random generator according to the model $p_m(\mathbf{y})$ of the PDF of $\mathbf{y}$. In our evaluation, we have used $N_\nu = 10^6$ synthetic vectors.

In the simulations we found a differential entropy of $\mathbf{y}$ of about $h(\mathbf{y}) \approx -0.105$ nats. The resulting bound on the achievable mean log spectral distortion $\bar{d}_{\text{LSD}}$ in dependence of the available mutual information $I(\mathbf{x}; \mathbf{y})$ between the feature vector $\mathbf{x}$ of the bandlimited speech and the cepstral representation $\mathbf{y}$ of the missing frequency band is illustrated in Fig. 3.

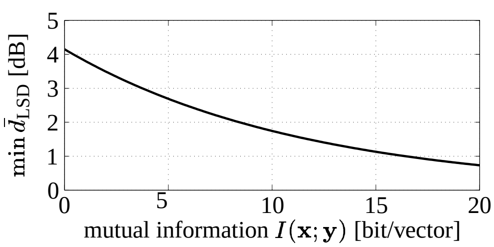

Fig. 3. Lower bound on the RMS log spectral distortion $\bar{d}_{\text{LSD}}$ for the bandwidth extension towards high frequencies.

In analogy to the estimation of the differential entropy, an approximation of the MI $I(\mathbf{x}; \mathbf{y})$ can be determined with help of a GMM $p_m(\mathbf{x}, \mathbf{y}) \approx p(\mathbf{x}, \mathbf{y})$ of the joint PDF of $\mathbf{x}$ and $\mathbf{y}$. The numerical evaluation is based on the integral definition of the MI [13], and it can be performed in the same manner as for the differential entropy with synthetic pairs of data vectors $\mathbf{x}_m$ and $\mathbf{y}_m$

$$I(\mathbf{x}; \mathbf{y}) \approx E_m \left\{ \log \frac{p_m(\mathbf{x}_m, \mathbf{y}_m)}{p_m(\mathbf{x}_m)\, p_m(\mathbf{y}_m)} \right\}. \tag{20}$$

The results of estimates of the MI for some conceivable feature vectors $\mathbf{x}$ of the bandlimited speech signal are listed in Tab. 4.

It must be emphasized that the accuracy of the approximated values for the entropy $h(\mathbf{y})$ and MI $I(\mathbf{x}; \mathbf{y})$ as presented in this section are limited by the accuracy of the constituting GMMs. Hence, the figures given in the example of this section are by no means the final answer.

feature vector $\mathbf{x}$	dim $\mathbf{x}$	$I(\mathbf{y}; \mathbf{x})$ [bit/vector]
mel freq. cepstral coeff. (MFCC)	10	2.3325
auto-correlation function	10	2.6089
LPC coefficients	10	2.3054
cepstrum (LPC-derived)	10	2.2401

Table 1. Estimates of the MI $I(\mathbf{x}; \mathbf{y})$ for the BWE of the upper frequency band with different features of the bandlimited speech.

5. CONCLUSIONS

In this paper we have presented a lower bound on the mean log spectral distortion of the spectral envelope in the missing frequency band as achievable by *any* memory-less BWE algorithm. Since the tightness of the bound has not been proven yet, it is not known whether effective estimators exist. Also, it is in the nature of the information theoretic bound that it does not point out a particular strategy to design optimal estimators.

From the fact that the maximum achievable performance of a BWE system depends on the MI $I(\mathbf{x}; \mathbf{y})$, we can conclude that it is advantageous to carefully select the elements of the utilized feature vector $\mathbf{x}$ such as to maximize the mutual information.

6. REFERENCES

[1] Y.M. Cheng, D. O'Shaughnessy, and P. Mermelstein, "Statistical Recovery of Wideband Speech from Narrowband Speech," *IEEE Trans. Speech and Audio Proc.*, vol. 2, no. 4, pp. 544–548, Oct. 1994.

[2] H. Carl, *Untersuchung verschiedener Methoden der Sprachkodierung und eine Anwendung zur Bandbreitenvergrößerung von Schmalband-Sprachsignalen*, Ph.D. thesis, Ruhr-Universität Bochum, 1994, (in German).

[3] C. Avendano, H. Hermansky, and E.A. Wan, "Beyond Nyquist: Towards the Recovery of Broad-Bandwidth Speech from Narrow-Bandwidth Speech," in *Proc. of European Conf. on Speech Communication and Technology (EUROSPEECH)*, Madrid, Sept. 1995.

[4] J. Epps and W.H. Holmes, "A New Technique for Wideband Enhancement of Coded Narrowband Speech," in *Proc. of IEEE Workshop on Speech Coding*, Porvoo, Finland, Sept. 1999.

[5] K.-Y. Park and H.S. Kim, "Narrowband to Wideband Conversion of Speech using GMM-based Transformation," in *Proc. of ICASSP*, Istanbul, June 2000, vol. 3, pp. 1847–1850.

[6] P. Jax and P. Vary, "Wideband Extension of Telephone Speech Using a Hidden Markov Model," in *Proc. of IEEE Workshop on Speech Coding*, Delavan, Wisconsin, Sept. 2000, pp. 133–135.

[7] M. Nilsson and W.B. Kleijn, "Avoiding Over-Estimation in Bandwidth Extension of Telephony Speech," in *Proc. of ICASSP*, Salt Lake City, May 2001.

[8] M. Nilsson, S.V. Andersen, and W.B. Kleijn, "On the Mutual Information Between Frequency Bands in Speech," in *Proc. of ICASSP*, Istanbul, June 2000, vol. 3, pp. 1327–1330.

[9] F. Nordén, T. Eriksson, and P. Hedelin, "An Information Theoretic Perspective on the Speech Spectrum Process," in *Proc. of IEEE Workshop on Speech Coding*, Delavan, Wisconsin, Sept. 2000.

[10] J.D. Markel and A.H. Gray, *Linear Prediction of Speech*, Springer-Verlag, Berlin, Heidelberg, New York, 1976.

[11] R. Hagen, "Spectral Quantization of Cepstral Coefficients," in *Proc. of ICASSP*, Adelaide, Apr. 1994, vol. 1, pp. 509–512.

[12] H.-P. Bernhard, "A Tight Upper Bound on the Gain of Linear and Nonlinear Predictors for Stationary Stochastic Processes," *IEEE Trans. Signal Proc.*, vol. 46, no. 11, pp. 2909–2917, Nov. 1998.

[13] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications, 1991.

[14] S.V. Vaseghi, *Advanced Signal Processing and Digital Noise Reduction*, Wiley, Teubner, 1996.

[15] P. Hedelin and J. Skoglund, "Vector Quantization Based on Gaussian Mixture Models," *IEEE Trans. Speech and Audio Proc.*, vol. 8, no. 4, pp. 385–401, July 2000.