AN INFORMATION THEORETIC VIEW ON ARTIFICIAL BANDWIDTH EXTENSION IN NOISY ENVIRONMENTS

Thomas Esch and Peter Vary

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

ABSTRACT

Artificial Bandwidth Extension (ABWE) exploits spectral dependencies of speech signals and aims at recovering missing high frequency components if only the narrowband speech signal is available. This contribution provides an information theoretic view on ABWE when used in noisy conditions. Based on the results of [1], a performance bound of ABWE is formulated if the narrowband signal is disturbed by additive noise. The performance bound is evaluated using real entropy measurements and the influence of noise suppression prior to ABWE is investigated.

Index Terms— Artificial bandwidth extension, mutual information, noise reduction

1. INTRODUCTION

Various wideband (50 Hz – 7 kHz) speech codecs have been developed in the past in order to increase the quality of today's telephone speech and are about to be introduced in current mobile networks. Depending on the infrastructure and the terminals that are available, several possibilities exist how to set up a wideband connection. The use of dedicated wideband codecs achieves the highest speech quality but requires a modification of the whole communication system. In contrast, when using Artificial Bandwidth Extension (ABWE), only the decoder has to be changed. This technique exploits spectral dependencies of speech signals in order to recover missing high frequency components by utilizing only the narrowband (50 Hz – 3.4 kHz) speech signal, e.g., [2]. The use of ABWE techniques is fully compatible with existing narrowband speech communication systems. This is important as the change of the current bandwidth limitation in public telephony systems will not happen abruptly. Although the concept of ABWE does not achieve the full quality of true wideband coding, it can be used to improve the acceptance by the user while achieving a smooth transition between narrowband and wideband speech coding.

In the derivation and training phase of an ABWE system, clean speech signals are available and can be applied to the system. However, if the algorithm is used in practical speech communication systems, the quality of the narrowband signal is often impaired due to background noise. In this case, the performance of ABWE significantly degrades. This paper investigates ABWE in noisy environments from an information theoretic point of view.

The mutual information between frequency bands in clean speech is examined, e.g., in [3] and [1]. Following the approach of [4], an upper bound on the quality of ABWE techniques is derived in [1] for the case that the clean narrowband signal is available. As it is likely that the recorded signal is disturbed by ambient noise in a realistic scenario, the derivation of [1] is extended in this contribution. A theoretic bound on the performance of ABWE is formulated if the narrowband signal is disturbed by Additive White Gaussian Noise (AWGN). Afterwards, real entropy measurements demonstrate the existence of spectral dependencies between low and high frequencies for a broad SNR range. Moreover, the influence of noise reduction techniques applied to the disturbed narrowband signal prior to ABWE is analyzed.

The remainder of this paper is organized as follows. In Sec. 2, a brief overview of the considered ABWE system is given. The performance bound of the system is derived in Sec. 3 incorporating a possibly disturbed narrowband signal. Real entropy measurements are presented in Sec. 4 and finally, conclusions are drawn in Sec. 5.

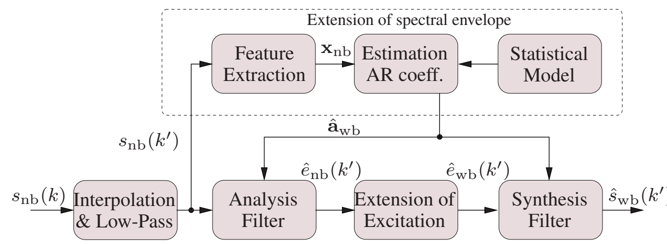

Fig. 1. Block diagram of the artificial bandwidth extension system for a noise-free input signal s_{nb}.

2. SYSTEM OVERVIEW

In Fig. 1, a simplified block diagram of the analyzed ABWE system is depicted which is proposed in [2]. At first, the sampling frequency of the narrowband input speech signal $s_{\mathrm{nb}}(k)$ is increased from $f_{\mathrm{s}} = 8$ kHz to $f_{\mathrm{s}} = 16$ kHz by interpolation and subsequent low-pass filtering. From now on all further steps are applied to the upsampled narrowband signal $s_{\mathrm{nb}}(k')$ at the sampling frequency $f_{\mathrm{s}} = 16$ kHz where k and k' denote the time instances in the respective domains.

Based on $s_{\mathrm{nb}}(k')$ the spectral envelope of the narrowband signal is extended in the upper part of the block diagram. An estimate $\hat{\mathbf{a}}_{\mathrm{wb}}$ of the feature vector $\mathbf{a}_{\mathrm{wb}}$ representing the spectral envelope of the wideband signal $s_{\mathrm{wb}}(k')$ is calculated using, e.g., a Hidden Markov Model (HMM) as proposed in [2]. The vector $\hat{\mathbf{a}}_{\mathrm{wb}}$ consists, e.g., of Autoregressive (AR) coefficients and is determined by exploiting information from an observation vector $\mathbf{x}_{\mathrm{nb}}$ as well as a priori knowledge provided by a pre-trained statistical model. Usually the vector $\mathbf{x}_{\mathrm{nb}}$ itself also contains information about the spectral envelope of the input signal $s_{\mathrm{nb}}(k')$.

The estimated vector $\hat{\mathbf{a}}_{\mathrm{wb}}$ is used to form a Finite Impulse Response (FIR) analysis filter which is applied to the input signal $s_{\mathrm{nb}}(k')$ in order to obtain an estimate of the bandlimited narrowband excitation signal $\hat{e}_{\mathrm{nb}}(k')$. In the next step, the missing extension band frequencies in the excitation signal are determined. As the human ear is relatively insensitive to variations of the spectral fine structure at high frequencies, the procedure can be implemented quite efficiently. Different approaches can be found, e.g., in [2].

Finally, the estimated wideband excitation signal $\hat{e}_{\mathrm{wb}}(k')$ is combined with the envelope features of the vector $\hat{\mathbf{a}}_{\mathrm{wb}}$ using a synthesis filter which is inverse to the applied analysis filter. The resulting signal $\hat{s}_{\mathrm{wb}}(k')$ provides an estimate of the wideband speech signal and exhibits transparency with respect to the narrowband input signal $s_{\mathrm{nb}}(k')$.

3. PERFORMANCE BOUND IN NOISY ENVIRONMENTS

In this section, the spectral dependencies between the disturbed narrowband and the clean extension band speech signal (3.4 kHz – 7 kHz) are analyzed based on the observation vector $\mathbf{x}_{\mathrm{nb}}$ and the feature vector $\mathbf{a}_{\mathrm{eb}}$ representing the spectral envelope of the extension band only, i.e., $\mathbf{a}_{\mathrm{wb}} = f(\mathbf{x}_{\mathrm{nb}}, \mathbf{a}_{\mathrm{eb}})$.

In order to determine a performance bound for the quality of ABWE in noisy environments, the derivation of the bound in [1] is extended considering the signal-flow model shown in Fig. 2a). The b_{nb}-dimensional observation vector $\mathbf{x}_{\mathrm{nb}}$ is assumed to be degraded by the additive noise vector $\mathbf{n}_{\mathrm{nb}}$ of dimension b_{nb} as well. Afterwards, ABWE is performed based on the resulting noisy observation vector $\mathbf{x}_{\mathrm{dnb}}$. Therefore, ABWE is described by the function $f_{\mathrm{bwe}}(\cdot)$, yielding the estimated feature vector $\hat{\mathbf{a}}_{\mathrm{eb}}$ of dimension b_{eb} according to:

$$\hat{\mathbf{a}}_{\mathrm{eb}} = f_{\mathrm{bwe}}\left(\mathbf{x}_{\mathrm{dnb}}\right). \tag{1}$$

The estimation error of the ABWE process is defined as b_{eb}-dimensional vector $\mathbf{n}_{\mathrm{eb}}$ and states the difference between $\mathbf{a}_{\mathrm{eb}}$ and $\hat{\mathbf{a}}_{\mathrm{eb}}$:

$$\mathbf{n}_{\mathrm{eb}} = \mathbf{a}_{\mathrm{eb}} - \hat{\mathbf{a}}_{\mathrm{eb}}. \tag{2}$$

Following the concept of [1], it is assumed that the ABWE system uses a memoryless estimator $f_{\mathrm{bwe}}(\cdot)$ which is not relying on information from previous or subsequent frames. Moreover, the process of the disturbances by $\mathbf{n}_{\mathrm{nb}}$ and $\mathbf{n}_{\mathrm{eb}}$ including the ABWE estimation is modeled by two independent memoryless, additive noisy channels. The resulting information theoretic dependencies between $\mathbf{x}_{\mathrm{nb}}$, $\mathbf{x}_{\mathrm{dnb}}$ and $\mathbf{a}_{\mathrm{eb}}$ are depicted in Fig. 2b) based on (conditional) differential entropies $h(\cdot)$. The mutual information $I(\mathbf{x}_{\mathrm{nb}};\mathbf{a}_{\mathrm{eb}})$ describes the linear and non-linear dependencies between $\mathbf{x}_{\mathrm{nb}}$ and $\mathbf{a}_{\mathrm{eb}}$, i.e., a high mutual information between the two vectors is desirable in order to obtain a good estimate of $\mathbf{a}_{\mathrm{eb}}$. As shown in Fig. 2b), the mutual information $I(\mathbf{x}_{\mathrm{nb}};\mathbf{a}_{\mathrm{eb}})$ can be expressed as:

$$\begin{aligned} I(\mathbf{x}_{\mathrm{nb}};\mathbf{a}_{\mathrm{eb}}) &= I(\mathbf{x}_{\mathrm{dnb}};\mathbf{a}_{\mathrm{eb}}) - h(\mathbf{x}_{\mathrm{dnb}}|\mathbf{x}_{\mathrm{nb}}) \\ &= h(\mathbf{a}_{\mathrm{eb}}) - h(\mathbf{a}_{\mathrm{eb}}|\mathbf{x}_{\mathrm{dnb}}) - h(\mathbf{x}_{\mathrm{dnb}}|\mathbf{x}_{\mathrm{nb}}), \end{aligned} \tag{3}$$

where $I(\mathbf{x}_{\mathrm{dnb}};\mathbf{a}_{\mathrm{eb}})$ represents the mutual information between $\mathbf{x}_{\mathrm{dnb}}$ and $\mathbf{a}_{\mathrm{eb}}$, $h(\mathbf{a}_{\mathrm{eb}})$ the differential entropy of $\mathbf{a}_{\mathrm{eb}}$, $h(\mathbf{a}_{\mathrm{eb}}|\mathbf{x}_{\mathrm{dnb}})$ the conditional differential entropy of $\mathbf{a}_{\mathrm{eb}}$ when $\mathbf{x}_{\mathrm{dnb}}$ is given and $h(\mathbf{x}_{\mathrm{dnb}}|\mathbf{x}_{\mathrm{nb}})$ the conditional differential entropy of $\mathbf{x}_{\mathrm{dnb}}$ when $\mathbf{x}_{\mathrm{nb}}$ is known. The latter can further be simplified [5]:

$$\begin{aligned} h(\mathbf{x}_{\mathrm{dnb}}|\mathbf{x}_{\mathrm{nb}}) &= h(\mathbf{x}_{\mathrm{nb}} + \mathbf{n}_{\mathrm{nb}}|\mathbf{x}_{\mathrm{nb}}) \\ &= h(\mathbf{n}_{\mathrm{nb}}|\mathbf{x}_{\mathrm{nb}}) \\ &= h(\mathbf{n}_{\mathrm{nb}}). \end{aligned} \tag{4}$$

The disturbance of the observation vector $\mathbf{x}_{\mathrm{nb}}$ by $\mathbf{n}_{\mathrm{nb}}$ can be interpreted as transmission over b_{nb} different channels in parallel. Assuming a fixed variance for the observation vector, an upper bound

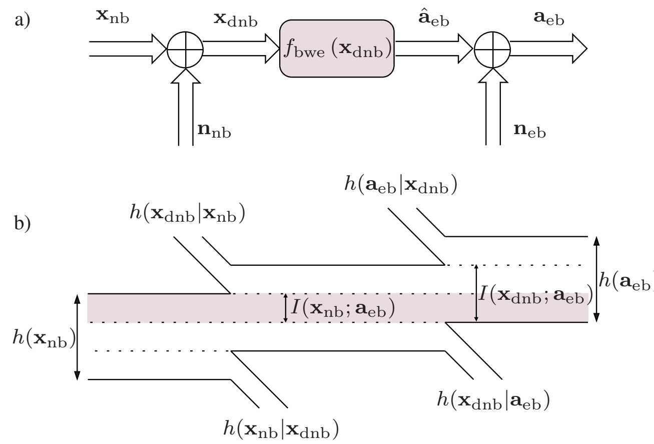

Fig. 2. Artificial bandwidth extension in noisy environments based on two memoryless channels: a) signal-flow model and b) mutual information between $\mathbf{x}_{\mathrm{nb}}$ and $\mathbf{a}_{\mathrm{eb}}$.

for the entropy $h(\mathbf{n}_{\mathrm{nb}})$ is given for the case that all channels are statistically independent$^{(a)}$ AWGN$^{(b)}$ channels with same variances$^{(c)}$ $\sigma^2_{n_{\mathrm{nb}},j} = \sigma^2_{n_{\mathrm{nb}}}$ for $0 \leq j < b_{\mathrm{nb}}$ [5]:

$$\begin{aligned} h(\mathbf{n}_{\mathrm{nb}}) &= h(n_{\mathrm{nb},0},\, n_{\mathrm{nb},1},\, \ldots,\, n_{\mathrm{nb},b_{\mathrm{nb}}-1}) \\ &\overset{(a)}{=} \sum_{j=0}^{b_{\mathrm{nb}}-1} h(n_{\mathrm{nb},j}) \\ &\overset{(b)}{\leq} \sum_{j=0}^{b_{\mathrm{nb}}-1} \frac{1}{2}\log_2(2\pi \mathrm{e}\sigma^2_{n_{\mathrm{nb}},j}) \\ &\overset{(c)}{=} b_{\mathrm{nb}}\log_2\left(\sqrt{2\pi \mathrm{e}\sigma^2_{n_{\mathrm{nb}}}}\right). \end{aligned} \tag{5}$$

The variable e states the Euler number and the unit of the (differential) entropies and mutual information is *bits/vector*. Under the assumption that the observation vector $\mathbf{x}_{\mathrm{nb}}$ consists of the first b_{nb} *cepstral coefficients*[1] $c_{\mathrm{nb},j}$ of the clean narrowband signal $s_{\mathrm{nb}}(k)$, i.e.:

$$x_{\mathrm{nb,j}} = \begin{cases} \frac{1}{\sqrt{2}}c_{\mathrm{nb},0} & \text{for } j = 0, \\ c_{\mathrm{nb},j} & \text{for } 1 \leq j < b_{\mathrm{nb}}, \end{cases} \tag{6}$$

the noise vector $\mathbf{n}_{\mathrm{nb}}$ is related to the Log Spectral Distortion (LSD) between the spectral envelopes of the clean narrowband signal and the disturbed narrowband signal, represented by the vectors $\mathbf{x}_{\mathrm{nb}}$ and $\mathbf{x}_{\mathrm{dnb}}$, respectively. The LSD measure $d^{\mathrm{LSD}}_{\mathrm{nb}}$ correlates well with the subjective speech quality and is defined as [1]:

$$\begin{aligned} d^{\mathrm{LSD}}_{\mathrm{nb}} =& \frac{\sqrt{2}\cdot 10}{\log_{\mathrm{e}}(10)}\sqrt{\mathbb{E}\Bigg\{\frac{1}{2}\left(c_{\mathrm{nb},0} - c_{\mathrm{dnb},0}\right)^2 \ldots} \\ & \overline{+\sum_{j=1}^{\infty}\left(c_{\mathrm{nb},j} - c_{\mathrm{dnb},j}\right)^2\Bigg\}}, \end{aligned} \tag{7}$$

where $\mathbb{E}\{\cdot\}$ represents the expectation operator and $c_{\mathrm{dnb},j}$ the j-th cepstral coefficient of the disturbed narrowband signal included in $\mathbf{x}_{\mathrm{dnb}}$. The unit of $d^{\mathrm{LSD}}_{\mathrm{nb}}$ is dB. A lower bound for $d^{\mathrm{LSD}}_{\mathrm{nb}}$ is given by:

$$d^{\mathrm{LSD}}_{\mathrm{nb}} \geq \frac{\sqrt{2}\cdot 10}{\log_{\mathrm{e}}(10)}\sqrt{\mathbb{E}\left\{|\mathbf{n}_{\mathrm{nb}}|^2\right\}} \geq \frac{\sqrt{2}\cdot 10}{\log_{\mathrm{e}}(10)}\sqrt{b_{\mathrm{nb}}\cdot\sigma^2_{n_{\mathrm{nb}}}}, \tag{8}$$

[1] It is shown in [6] that cepstral coefficients provide high information on the spectral envelope of the extension band.

which can be rearranged to:

$$\sqrt{\sigma^2_{n_{\rm nb}}} \leq \frac{d^{\rm LSD}_{\rm nb} \cdot \log_{\rm e}(10)}{\sqrt{2} \cdot 10 \cdot \sqrt{b_{\rm nb}}}. \tag{9}$$

Using Ineqs. 9 and 5 and Eq. 4, the mutual information $I(\mathbf{x}_{\rm nb}; \mathbf{a}_{\rm eb})$ in Eq. 3 is bounded by:

$$I(\mathbf{x}_{\rm nb}; \mathbf{a}_{\rm eb}) \geq \underbrace{h(\mathbf{a}_{\rm eb}) - h(\mathbf{a}_{\rm eb}|\mathbf{x}_{\rm dnb})}_{I(\mathbf{x}_{\rm dnb};\mathbf{a}_{\rm eb})} - b_{\rm nb} \log_2\left(\frac{\sqrt{\pi \rm e}\log_{\rm e}(10)}{10 \cdot \sqrt{b_{\rm nb}}} d^{\rm LSD}_{\rm nb}\right). \tag{10}$$

The mutual information $I(\mathbf{x}_{\rm dnb}; \mathbf{a}_{\rm eb})$ incorporates estimation errors of the artificial bandwidth extension which are represented in Fig. 2 by $\mathbf{n}_{\rm eb}$. An expression for $I(\mathbf{x}_{\rm dnb}; \mathbf{a}_{\rm eb})$ is derived in [1] using very similar calculus as above. Assuming that the upper frequency band is represented in $\mathbf{a}_{\rm eb}$ by cepstral coefficients of the extension band as well, the conditional entropy $h(\mathbf{a}_{\rm eb}|\mathbf{x}_{\rm dnb})$ is also bounded by the LSD $d^{\rm LSD}_{\rm eb}$ of the extension band according to:

$$h(\mathbf{a}_{\rm eb}|\mathbf{x}_{\rm dnb}) = h(\mathbf{n}_{\rm eb}) \leq b_{\rm eb} \log_2\left(\frac{\sqrt{\pi \rm e}\log_{\rm e}(10)}{10 \cdot \sqrt{b_{\rm eb}}} d^{\rm LSD}_{\rm eb}\right), \tag{11}$$

finally leading to the following lower bound for the mutual information between $\mathbf{x}_{\rm nb}$ and $\mathbf{a}_{\rm eb}$:

$$I(\mathbf{x}_{\rm nb}; \mathbf{a}_{\rm eb}) \geq h(\mathbf{a}_{\rm eb}) - b_{\rm eb} \log_2\left(\frac{\sqrt{\pi \rm e}\log_{\rm e}(10)}{10 \cdot \sqrt{b_{\rm eb}}} d^{\rm LSD}_{\rm eb}\right) - b_{\rm nb} \log_2\left(\frac{\sqrt{\pi \rm e}\log_{\rm e}(10)}{10 \cdot \sqrt{b_{\rm nb}}} d^{\rm LSD}_{\rm nb}\right). \tag{12}$$

Thereby, estimation errors occurring in the ABWE process are considered in the first subtrahend and the information loss due to disturbance of the narrowband signal is expressed by the second subtrahend. Knowing the distortion caused by a specific ABWE estimator as well as the degradation of the narrowband signal, the mutual information which is at least included in $\mathbf{x}_{\rm nb}$ and $\mathbf{a}_{\rm eb}$ is given by Ineq. 12. In the following, two scenarios are considered in which ABWE in noisy environments is analyzed based on real entropy measurements.

4. RESULTS

In this section, the performance bound is applied to real speech data and the influence of narrowband noise reduction prior to ABWE is investigated. Therefore, the noise suppression system proposed in [7] is applied to the disturbed narrowband signal yielding the narrowband speech estimate. In order to compare the cases before and after noise suppression, the observation vector $\mathbf{x}_{\rm dnb}$ is extracted from both, the noisy narrowband signal and the enhanced narrowband signal after noise reduction.

Two scenarios characterized by the choice of parameters used for the observation vectors $\mathbf{x}_{\rm nb}$ and $\mathbf{x}_{\rm dnb}$ as well as the feature vector $\mathbf{a}_{\rm eb}$ are considered. In Scenario I, the theoretical bound of Ineq. 12 is analyzed according to the measured differential entropy $h(\mathbf{a}^{\rm I}_{\rm eb})$ and the LSD measures $d^{\rm LSD}_{\rm nb}$ and $d^{\rm LSD}_{\rm eb}$. With respect to the derivation of this lower bound, cepstral coefficients are chosen as parameters for the narrowband as well as for the extension band according to:

Scenario I	Parameter(s)	Dimension
Observ. vector $\mathbf{x}^{\rm I}_{\rm nb}$	Cepstral coefficients	$b_{\rm nb} = 10$
Observ. vector $\mathbf{x}^{\rm I}_{\rm dnb}$	Cepstral coefficients	$b_{\rm nb} = 10$
Feature vector $\mathbf{a}^{\rm I}_{\rm eb}$	Cepstral coefficients	$b_{\rm eb} = 10$

Scenario II considers the wideband noise suppression system which is proposed in [8]. The system uses ABWE techniques in order to improve the spectral estimation process in the extension band. Therefore, features from the processed (enhanced) narrowband signal are extracted and used to estimate subband energies of the extension band. These energy estimates support a conventional noise reduction technique in the extension band. In Scenario II, the effective mutual information $I(\mathbf{x}^{\rm II}_{\rm dnb}; \mathbf{a}^{\rm II}_{\rm eb})$ which can be achieved in [8] by ABWE is investigated and its dependency on the input SNR is shown. In order to get results which can be transferred to the actual system in [8], the same parameters and dimensions as used in [8] are applied for the observation and feature vectors, namely Mel Frequency Cepstral Coefficients (MFCCs), the Zero Crossing Rate (ZCR) as well as subband energies as follows:

Scenario II	Parameter(s)	Dimension
Observ. vector $\mathbf{x}^{\rm II}_{\rm nb}$	MFCCs + ZCR	$b_{\rm nb} = 14$
Observ. vector $\mathbf{x}^{\rm II}_{\rm dnb}$	MFCCs + ZCR	$b_{\rm nb} = 14$
Feature vector $\mathbf{a}^{\rm II}_{\rm eb}$	Subband energies	$b_{\rm eb} = 12$

The measurements of $h(\mathbf{a}^{\rm I}_{\rm eb})$, $h(\mathbf{a}^{\rm II}_{\rm eb})$ and $I(\mathbf{x}^{\rm II}_{\rm dnb}; \mathbf{a}^{\rm II}_{\rm eb})$ are carried out by using the well-known *k-nearest neighbor* algorithm [9] to estimate the required probability density functions [5]. The algorithm is data efficient, adaptive and achieves minimal bias. The number k here decides how many neighbors influence the final classification. In the sequel, the results for the two scenarios are presented using $k = 1$.

For the evaluation, about 16 minutes of speech taken from the NTT speech database (sampling frequency $f_{\rm s}$ =16 kHz) are disturbed by AWGN at different SNR values. The wideband signals are split into the respective narrowband and extension band part before the signals are downsampled by a factor of 2. In the simulation setup, the speech and noise signals are both available and can therefore be filtered and downsampled separately. Hence, clean and noisy versions of narrowband as well as extension band signals are accessible. Based on these signals, the observation vectors $\mathbf{x}_{\rm nb}$ and $\mathbf{x}_{\rm dnb}$ as well as the feature vector $\mathbf{a}_{\rm eb}$ are extracted using 20 ms non-overlapping frames contributing to speech activity.

4.1. Scenario I (Narrowband Noise Reduction)

In order to apply Ineq. 12, this scenario considers the use of cepstral coefficients as parameters within the observation and feature vectors. The LSD $d^{\rm LSD}_{\rm nb}$ of the narrowband as well as the differential entropy $h(\mathbf{a}^{\rm I}_{\rm eb})$ are determined from real data. The narrowband speech signal $s_{\rm nb}(k)$ is disturbed by AWGN at input SNR values of -15 dB, 0 dB and 15 dB. It has to be mentioned that the disturbance of $s_{\rm nb}(k)$ by AWGN does not necessarily mean that the elements of the vector $\mathbf{n}_{\rm nb}$ are Gaussian distributed as well. However, in any case Ineq. 12 is valid as the entropy of a scalar signal with variance $\sigma^2_{n_{\rm nb}}$ is upper bounded by the entropy of a normally distributed variable with the same variance [5]. Table 1 shows the averaged LSD values $d^{\rm LSD}_{\rm nb}$ measured before and after noise suppression is applied. Moreover, the differential entropy $h(\mathbf{a}^{\rm I}_{\rm eb})$ is measured and yields

Log. spectral dist. $d^{\rm LSD}_{\rm nb}$	Input SNR		
	-15 dB	0 dB	15 dB
Without noise reduction	4.627 dB	2.529 dB	1.083 dB
With noise reduction	2.047 dB	1.115 dB	0.944 dB

Table 1. Averaged log spectral distortion $d^{\rm LSD}_{\rm nb}$ of disturbed narrowband signal measured before and after noise suppression.

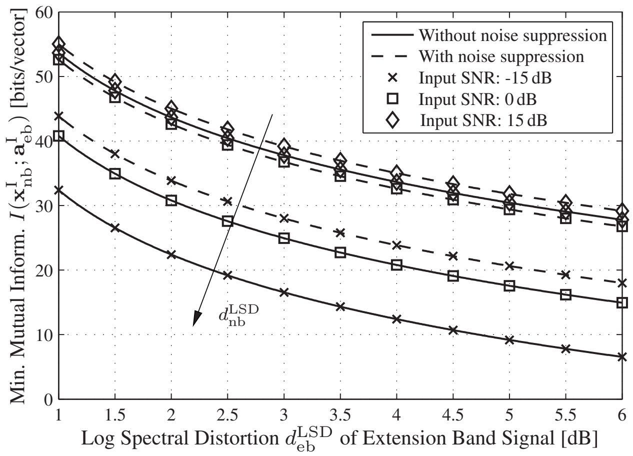

Fig. 3. Theoretical lower bound according to Ineq. 12 for the mutual information $I(\mathbf{x}_{\mathrm{nb}}^{\mathrm{I}};\mathbf{a}_{\mathrm{eb}}^{\mathrm{I}})$ between extension and narrowband dependent on the log spectral distortions $d_{\mathrm{nb}}^{\mathrm{LSD}}$ and $d_{\mathrm{eb}}^{\mathrm{LSD}}$.

$h(\mathbf{a}_{\mathrm{eb}}^{\mathrm{I}}) \approx 9.533$ bits/vector in this setup. Based on these measurements, Fig. 3 depicts the theoretical lower bounds for the mutual information $I(\mathbf{x}_{\mathrm{nb}}^{\mathrm{I}};\mathbf{a}_{\mathrm{eb}}^{\mathrm{I}})$ according to Ineq. 12 while varying $d_{\mathrm{eb}}^{\mathrm{LSD}}$ from 1 dB to 6 dB. The figure clearly shows the dependency on the two distortion measures: the lower $d_{\mathrm{nb}}^{\mathrm{LSD}}$ or $d_{\mathrm{eb}}^{\mathrm{LSD}}$ the higher the theoretical bound and vice versa. In addition, the advantage of applying noise suppression prior to ABWE can be seen. According to Tab. 1, noise reduction achieves a reduction of the LSD measure $d_{\mathrm{nb}}^{\mathrm{LSD}}$ leading to a higher bound $I(\mathbf{x}_{\mathrm{nb}}^{\mathrm{I}};\mathbf{a}_{\mathrm{eb}}^{\mathrm{I}})$ compared to the case when no noise reduction is applied. Figure 3 illustrates this behavior for the three investigated SNR values where the discrepancy is especially high for -15 dB input SNR.

4.2. Scenario II (Noise Reduction Supported by ABWE)

In this scenario, the ABWE which is used in [8] is considered. The system exploits mutual information between the enhanced narrowband signal and the extension band speech signal by ABWE techniques. In the following, the mutual information $I(\mathbf{x}_{\mathrm{dnb}}^{\mathrm{II}};\mathbf{a}_{\mathrm{eb}}^{\mathrm{II}})$ between the disturbed observation vector $\mathbf{x}_{\mathrm{dnb}}^{\mathrm{II}}$, which is available in the system, and the feature vector $\mathbf{a}_{\mathrm{eb}}^{\mathrm{II}}$ is measured. The measurement is conducted in dependence on the input SNR and whether noise suppression is applied to the narrowband signal or not. For the evaluation purpose, the feature vector $\mathbf{a}_{\mathrm{eb}}^{\mathrm{II}}$ is extracted directly from the extension band of the original wideband signal and not from the estimated signal after ABWE. The results are shown in Fig. 4 for input SNR values varying from -20 dB to 35 dB. It can be seen that the mutual information is continuously increasing with the input SNR and finally converges to $I(\mathbf{x}_{\mathrm{nb}}^{\mathrm{II}};\mathbf{a}_{\mathrm{eb}}^{\mathrm{II}})$, i.e., the case where $\mathbf{x}_{\mathrm{dnb}}^{\mathrm{II}} = \mathbf{x}_{\mathrm{nb}}^{\mathrm{II}}$ or $\mathbf{n}_{\mathrm{nb}} = \mathbf{0}$. In addition, the figure also motivates the application of noise suppression before ABWE is performed: the mutual information with prior noise reduction is significantly higher than without noise reduction.

5. CONCLUSIONS

This paper covers artificial bandwidth extension in noisy environments from an information theoretic point of view. A performance bound for the quality of ABWE is derived assuming the narrowband speech signal to be disturbed by AWGN. The theoretical bound is

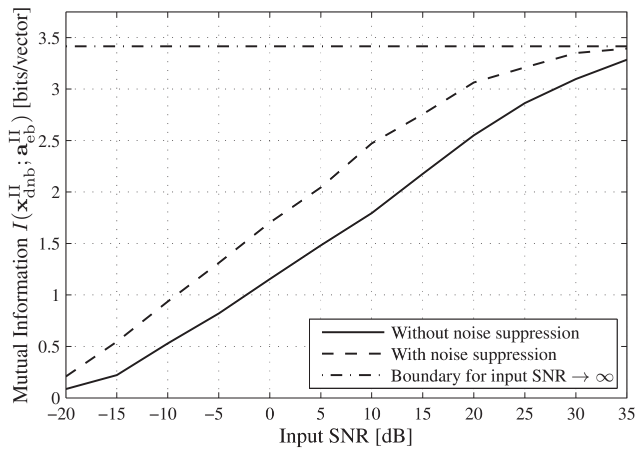

Fig. 4. Measured amount of mutual information $I(\mathbf{x}_{\mathrm{dnb}}^{\mathrm{II}};\mathbf{a}_{\mathrm{eb}}^{\mathrm{II}})$ between extension and narrowband as a function of the input SNR.

analyzed by real entropy measurements showing the existence of dependencies between narrowband and extension band even if the narrowband signal is severely disturbed by additive noise. In order to exploit this mutual information, the application of noise suppression to the narrowband signal is advantageous by all means before estimating the extension band parameters.

6. ACKNOWLEDGEMENTS

The authors would like to thank Bernd Geiser for software development and valuable contributions.

7. REFERENCES

[1] P. Jax and P. Vary, “An Upper Bound on the Quality of Artificial Bandwidth Extension of Narrowband Speech Signals,” in *Proc. of ICASSP*, Orlando, FL, USA, May 2002, pp. 237–240.

[2] P. Jax and P. Vary, “On Artificial Bandwidth Extension of Telephone Speech,” *Signal Processing*, vol. 83, no. 8, pp. 1707–1719, Aug. 2003.

[3] M. Nilsson, H. Gustafsson, S. V Andersen, and B. W. Kleijn, “Gaussian Mixture Model Based Mutual Information Estimation Between Frequency Bands in Speech,” in *Proc. of ICASSP*, Orlando, FL, USA, Apr. 2002.

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

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Chichester, UK, 2006.

[6] P. Jax and P. Vary, “Feature Selection for Improved Bandwidth Extension of Speech Signals,” in *Proc. of ICASSP*, Montreal, Canada, May 2004, pp. 697–700.

[7] T. Esch and P. Vary, “Model-Based Speech Enhancement Using SNR Dependent MMSE Estimation,” in *Proc. of ICASSP*, Prague, Czech Republic, May 2011.

[8] T. Esch, F. Heese, B. Geiser, and P. Vary, “Wideband Noise Suppression Supported by Artificial Bandwidth Extension Techniques,” in *Proc. of ICASSP*, Dallas, TX, USA, Mar. 2010, pp. 4790–4793.

[9] L. F. Kozachenko and N. N. Leonenko, “A Statistical Estimate for the Entropy of a Random Vector,” *Problems of Information Transmission*, vol. 23, no. 2, pp. 95–101, 1987.