

ARTIFICIAL BANDWIDTH EXTENSION OF SPEECH SIGNALS USING MMSE ESTIMATION BASED ON A HIDDEN MARKOV MODEL

Peter Jax and Peter Vary

Institute of Communication Systems and Data Processing (IND),
Aachen University (RWTH), 52056 Aachen, Germany
E-mail: jax@ind.rwth-aachen.de

ABSTRACT

We present an algorithm to derive 7 kHz wideband speech from narrowband “telephone speech”. A statistical approach is used that is based on a *Hidden Markov Model* (HMM) of the speech production process. A new method for the estimation of the wideband spectral envelope is proposed, using nonlinear state-specific techniques to minimize a mean square error criterion. In contrast to common *memoryless* estimation methods, additional information from adjacent signal frames can be exploited by utilizing the HMM. A consistent advantage of the new estimation rule is obtained compared to previously published HMM-based hard or soft classification methods.

1. INTRODUCTION

In current public telephone systems the bandwidth of the transmitted speech is typically limited due to legacies of the old analogue telephone system to a frequency range of up to 3.4 kHz. This bandwidth limitation causes the characteristic sound of “telephone speech”. Listening experiments have shown that an enlarged frequency bandwidth of speech signals contributes significantly to the perceived speech quality, e.g. [1]. This fact is reflected by ongoing standardizations of wideband speech codecs with a frequency range up to 7 kHz, e.g. [2]. However, true wideband speech communication requires a modification of the transmission link by enhanced speech codecs. Hence, for economical reasons there will be mixed telephone networks for a long transitional period, comprising both narrowband and wideband codecs.

An alternative approach towards a larger perceived (audio) bandwidth is *artificial bandwidth extension* (BWE): this technique aims at the recovery of missing low and/or high frequency components of the speech signal utilizing only the available narrowband speech at the receiving side of the transmission link. Bandwidth extension is feasible due to mutual dependencies in the frequency bands of speech signals [3, 4]. One possible concept to explore these redundancies is based on a linear source-filter model of the speech production process, as proposed in [5, 6]: The time-varying parameters of the source model are estimated from the narrowband speech. These parameters are then used in combination with the source model to estimate and add the missing frequency components.

In this paper, bandwidth extension towards high frequencies is treated: according to the frequency range of typical telephone speech, the narrowband (subscript nb) input signal $s_{nb}(k)$ is assumed to contain frequencies lower than 3.4 kHz. The extended frequency range between 3.4 and 8 kHz will be called *extension frequency band* (subscript eb) in the following. Wideband speech is marked by the subscript wb. Note that the results of this pa-

per can be applied to different BWE scenarios as well, e.g., to the BWE toward low frequencies [7].

2. BANDWIDTH EXTENSION ALGORITHM

The vast majority of the adaptive BWE algorithms published in literature to date is based on the well-known linear source-filter model of the speech production process: it is assumed that the human vocal tract (including glottis pulse shaping) 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 [5, 6]. Fig. 1 shows the signal flow of our BWE algorithm.

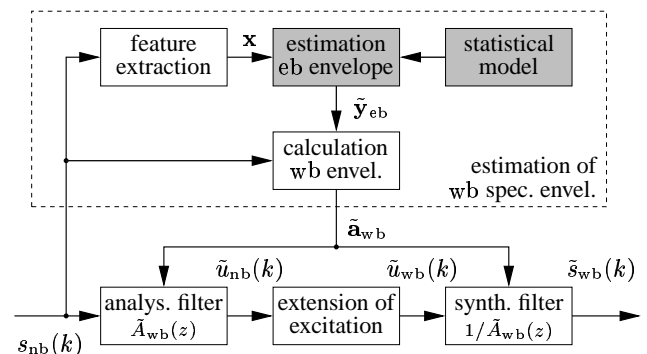


Fig. 1. Signal-flow of our BWE algorithm (from [8]). The focus of this paper is on the two sub-blocks that are shaded in gray.

In Fig. 1 it is assumed that the narrowband input signal $s_{nb}(k)$ is sampled at a sampling frequency that is sufficient to represent the extended wideband speech signal (e.g. $f_s = 16$ kHz). The processing is performed frame-by-frame with a frame-size of 20 ms. The frame index will be denoted by m in the following.

2.1. Estimation of the Wideband Spectral Envelope

The first step in the bandwidth extension algorithm is the estimation of the spectral envelope of the original wideband speech signal. In Fig. 1 this task is performed by the upper blocks. The resulting estimate \tilde{a}_{wb} resembles the coefficient set of the all-pole (vocal tract) filter of the source-filter model. It has been found that the estimated wideband spectral envelope is particularly important for the quality of the extended speech, e.g. [5].

For the estimation of $\tilde{\mathbf{a}}_{wb}$ several *memoryless* approaches have been proposed in literature, e.g., codebook based methods [5, 9, 10], linear mapping [11, 12], or statistical estimation [6, 13]. Our estimation approach differs from most of these methods mainly in the following two respects:

1. The estimation is based on a *hidden Markov model* (HMM) of the speech generation process. Thereby, our estimator can exploit information from previous signal frames to enhance the estimation quality [14, 4, 7, 8].
2. Before calculating the coefficients $\tilde{\mathbf{a}}_{wb}$, a cepstral vector $\tilde{\mathbf{y}}_{eb}$ (see [4, 7, 8]) is estimated, representing the shape and the gain of the spectral envelope of the *extension band only*. This estimate is combined with the narrowband signal frame in a short-term power spectrum domain. The (wideband) AR coefficients $\tilde{\mathbf{a}}_{wb}$ are obtained via the autocorrelation function using standard linear prediction analysis techniques [7, 8].

In Fig. 1 the core of the spectral envelope estimation within our BWE system is illustrated by the gray shading. The task of the marked sub-system is to estimate the spectral envelope of the extension band $\tilde{\mathbf{y}}_{eb}$ based on the observation of a feature vector \mathbf{x} that is extracted from the narrowband speech signal $s_{nb}(k)$. For this purpose a technique is used that is somewhat similar to pattern recognition and relies on a statistical model of the speech generation process. The statistical model is trained off-line using a training data base of true wideband speech signals.

A new algorithm for the estimation of the vector $\tilde{\mathbf{y}}_{eb}$ is proposed in Section 4. The underlying statistical model (hidden Markov model) is introduced in Section 3.

2.2. Extension of the Excitation Signal

By utilizing the estimated wideband AR filter coefficients $\tilde{\mathbf{a}}_{wb}$ for an FIR analysis filter $\tilde{A}_{wb}(z)$, operating on the narrowband input signal $s_{nb}(k)$, an estimate $\tilde{u}_{nb}(k)$ of the narrowband excitation signal is determined. Note that the frequency response of the analysis filter is the inverse of the frequency response of the vocal tract (synthesis) filter.

According to the assumed source-filter model of the speech production process, the excitation signal $\tilde{u}_{nb}(k)$ resembles noise (during unvoiced sounds) or a harmonic tone (during voiced sounds) that is spectrally flat in the frequency range covered by $s_{nb}(k)$. New excitation components in the extension band are derived by a spectral shift of the narrowband excitation $\tilde{u}_{nb}(k)$. This shift is obtained by a modulation of $\tilde{u}_{nb}(k)$ by a cosine signal with a frequency of 3.4 kHz [15, 16, 8]. The combination of the modulated excitation with $\tilde{u}_{nb}(k)$ yields the estimate $\tilde{u}_{wb}(k)$ of the wideband excitation signal.

The estimated wideband excitation signal $\tilde{u}_{wb}(k)$ is fed into the all-pole synthesis filter $1/\tilde{A}_{wb}(z)$ to synthesize the enhanced output speech signal $\hat{s}_{wb}(k)$. Because the excitation signal is not modified within the frequency range covered by $\tilde{u}_{nb}(k)$, and because the analysis filter and the synthesis filter are mutually inverse, the narrowband speech $s_{nb}(k)$ is contained transparently in the output signal $s_{wb}(k)$.

3. HIDDEN MARKOV MODEL

The vector estimation $\tilde{\mathbf{y}} = \tilde{\mathbf{y}}_{eb}$ representing the spectral envelope of the extension band is based on a *Hidden Markov Model* (HMM) of the speech generation process. In the sequel the subscript eb

of $\tilde{\mathbf{y}}_{eb}$ will be omitted for the sake of brevity. Each state \mathcal{S}_i of the HMM ($i = 1, \dots, N_S$) is assigned to a typical speech sound (frame of 20 ms). The HMM states are defined during the off-line training phase of the algorithm by the centroids of a *vector quantization* (VQ) of the spectral envelope representation \mathbf{y}_{eb} . The VQ codebook is the set of all centroids $\mathcal{C} = \{\hat{\mathbf{y}}_1, \dots, \hat{\mathbf{y}}_{N_S}\}$. Each state \mathcal{S}_i of the HMM is assigned to one entry $\hat{\mathbf{y}}_i$ of the VQ codebook such that the number of HMM states is the same as the number of codebook entries. It is implicitly assumed that the state may change from frame to frame.

3.1. First-Order Markov Chain

In the sequel, it will be assumed that the state sequence is governed by a first order Markov chain. Two sets of parameters are necessary to describe the properties of the state sequence:

- the probability of occurrence $P(\mathcal{S}_i)$ of the i -th state \mathcal{S}_i without taking into account the HMM state at the preceding or following frame instant, and
- the transition probabilities $P(\mathcal{S}_i(m+1)|\mathcal{S}_j(m))$ for consecutive frame instants.

The state probabilities $P(\mathcal{S}_i)$ and transition probabilities $P(\mathcal{S}_i(m+1)|\mathcal{S}_j(m))$ are calculated during the off-line training phase from wideband training data using the true state sequence derived via vector quantization of \mathbf{y}_{eb} .

3.2. State-Specific Gaussian Mixture Model

In our new estimation rule the state-specific mutual dependencies between the variables \mathbf{x} and \mathbf{y} shall be taken into account. Therefore, a model of the conditional joint *Probability Density Function* (PDF) $p(\mathbf{x}, \mathbf{y}|\mathcal{S}_i)$ is needed. For the definition of this PDF the two column vectors $\mathbf{x} = [x_1, \dots, x_b]^T$ and $\mathbf{y} = [y_1, \dots, y_d]^T$ are combined in the vector $\mathbf{z} = [\mathbf{x}^T, \mathbf{y}^T]^T$. The dimension of the vector \mathbf{z} is $\dim \mathbf{z} = b + d$, where $b = \dim \mathbf{x}$ and $d = \dim \mathbf{y}$.

Since both \mathbf{x} and \mathbf{y} are multi-dimensional vectors with continuous ranges of values, a vast amount of memory would be needed for modeling the joint PDFs by histograms. Instead, a parametric *Gaussian Mixture Model* (GMM) is utilized here:

$$p(\mathbf{z}|\mathcal{S}_i) \approx \sum_{l=1}^L \rho_{\mathbf{z},il} \mathcal{N}(\mathbf{z}; \mu_{\mathbf{z},il}, \mathbf{V}_{\mathbf{z},il}) \quad (1)$$

with $0 \leq \rho_{\mathbf{z},il} \leq 1$ and $\sum_{l=1}^L \rho_{\mathbf{z},il} = 1$ [17]. The mixture components are the $(b + d)$ -dimensional Gaussian densities

$$\mathcal{N}(\mathbf{z}; \mu_{\mathbf{z},il}, \mathbf{V}_{\mathbf{z},il}) = \frac{e^{-\frac{1}{2}(\mathbf{z} - \mu_{\mathbf{z},il})^T \mathbf{V}_{\mathbf{z},il}^{-1} (\mathbf{z} - \mu_{\mathbf{z},il})}}{(2\pi)^{(b+d)/2} \sqrt{\det \mathbf{V}_{\mathbf{z},il}}} \quad (2)$$

The mean vectors $\mu_{\mathbf{z},il}$ and covariance matrices $\mathbf{V}_{\mathbf{z},il}$ can be split as follows:

$$\begin{aligned} \mu_{\mathbf{z},il} &= \begin{pmatrix} \mu_{\mathbf{x},il} \\ \mu_{\mathbf{y},il} \end{pmatrix}, \\ \mathbf{V}_{\mathbf{z},il} &= \begin{pmatrix} \mathbf{A}_{\mathbf{xx},il} & \mathbf{A}_{\mathbf{xy},il} \\ \mathbf{A}_{\mathbf{yx},il} & \mathbf{A}_{\mathbf{yy},il} \end{pmatrix}^{-1}. \end{aligned} \quad (3)$$

The GMM parameters $\rho_{\mathbf{z},il}$, $\mu_{\mathbf{z},il}$, and $\mathbf{V}_{\mathbf{z},il}$ (with $i = 1, \dots, N_S$ and $l = 1, \dots, L$) are obtained during the off-line training phase. For each HMM state \mathcal{S}_i the state-specific PDF $p(\mathbf{x}, \mathbf{y}|\mathcal{S}_i)$ is approximated using the *expectation-maximization* (EM) algorithm, e.g. [17].

4. ESTIMATION OF THE SPECTRAL ENVELOPE OF THE EXTENSION BAND

By our estimation rule a continuous estimation of the parameter vector \mathbf{y} (cepstral representation of the envelope of the extension band) shall be performed. The error criterion to be minimized is the mean square error

$$E\{\|\mathbf{y}(m) - \tilde{\mathbf{y}}(m)\|^2 | \mathbf{X}(m)\} \stackrel{!}{=} \min, \quad (4)$$

where $\mathbf{X}(m) = \{\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(m)\}$ is the set of all feature vectors that have been observed up to the m -th signal frame. The solution of Eq. (4) is the conditional expectation

$$\begin{aligned} \tilde{\mathbf{y}}(m) &= E\{\mathbf{y}(m) | \mathbf{X}(m)\} \\ &= \int_{\mathbb{R}^d} \mathbf{y}(m) p(\mathbf{y}(m) | \mathbf{X}(m)) d\mathbf{y}(m). \end{aligned} \quad (5)$$

Because we do not have a model of the conditional PDF $p(\mathbf{y}(m) | \mathbf{X}(m))$ in closed-form, this quantity is expressed indirectly via the states of the HMM

$$p(\mathbf{y}(m) | \mathbf{X}(m)) = \sum_{i=1}^{N_S} p(\mathbf{y}(m), \mathcal{S}_i(m) | \mathbf{X}(m)). \quad (6)$$

The PDF $p(\mathbf{y}(m), \mathcal{S}_i(m) | \mathbf{X}(m))$ can be split into two factors since, according to the definition of the HMM, the vectors $\mathbf{x}(m)$ and $\mathbf{y}(m)$ exclusively depend on the state $\mathcal{S}_i(m)$ in the m -th signal frame

$$\begin{aligned} p(\mathbf{y}(m), \mathcal{S}_i(m) | \mathbf{X}(m)) \\ = p(\mathbf{y}(m) | \mathcal{S}_i(m), \mathbf{x}(m)) P(\mathcal{S}_i(m) | \mathbf{X}(m)). \end{aligned} \quad (7)$$

Inserting (6) and (7) into Eq. (5) yields

$$\begin{aligned} \tilde{\mathbf{y}}(m) &= \sum_{i=1}^{N_S} P(\mathcal{S}_i(m) | \mathbf{X}(m)) \\ &\cdot \int_{\mathbb{R}^d} \mathbf{y}(m) p(\mathbf{y}(m) | \mathcal{S}_i(m), \mathbf{x}(m)) d\mathbf{y}(m). \end{aligned} \quad (8)$$

The integral at the right-hand side of Eq. (8) reflects the conditional expectation

$$\begin{aligned} \int_{\mathbb{R}^d} \mathbf{y}(m) p(\mathbf{y}(m) | \mathcal{S}_i(m), \mathbf{x}(m)) d\mathbf{y} \\ = E\{\mathbf{y}(m) | \mathcal{S}_i(m), \mathbf{x}(m)\}. \end{aligned} \quad (9)$$

The conditional expectation $E\{\mathbf{y}(m) | \mathcal{S}_i(m), \mathbf{x}(m)\}$ can be calculated from the parameters of the Gaussian mixture model of the joint PDF $p(\mathbf{x}(m), \mathbf{y}(m) | \mathcal{S}_i(m))$ as follows [18, 13]

$$\begin{aligned} E\{\mathbf{y}(m) | \mathcal{S}_i(m), \mathbf{x}(m)\} &= \sum_{l=1}^L \rho_{\mathbf{y} | \mathbf{x}, il} \cdot \\ &\cdot \left(\mu_{\mathbf{y}, il} - ((\mathbf{x}(m) - \mu_{\mathbf{x}, il})^T \mathbf{A}_{\mathbf{y}\mathbf{x}, il} \mathbf{A}_{\mathbf{y}\mathbf{y}, il}^{-1})^T \right), \end{aligned} \quad (10)$$

with

$$\rho_{\mathbf{y} | \mathbf{x}, il} = \frac{\rho_{il} \mathcal{N}(\mathbf{x}; \mu_{\mathbf{x}, il}, \mathbf{V}_{\mathbf{x}, il})}{\sum_{l=1}^L \rho_{il} \mathcal{N}(\mathbf{x}; \mu_{\mathbf{x}, il}, \mathbf{V}_{\mathbf{x}, il})}.$$

Inserting the conditional expectation into Eq. (8) leads to the *minimum mean square error* (MMSE) estimation rule

$$\begin{aligned} \tilde{\mathbf{y}}_{\text{MMSE}}(m) \\ = \sum_{i=1}^{N_S} E\{\mathbf{y}(m) | \mathcal{S}_i(m), \mathbf{x}(m)\} P(\mathcal{S}_i(m) | \mathbf{X}(m)), \end{aligned} \quad (11)$$

where (see [7, 8])

$$\begin{aligned} P(\mathcal{S}_i(m) | \mathbf{X}(m)) &= \frac{p(\mathcal{S}_i(m), \mathbf{X}(m))}{\sum_{i=1}^{N_S} p(\mathcal{S}_i(m), \mathbf{X}(m))}, \\ p(\mathcal{S}_i(m), \mathbf{X}(m)) &= p(\mathbf{x}(m) | \mathcal{S}_i(m)) p(\mathcal{S}_i(m), \mathbf{X}(m-1)), \\ p(\mathbf{x}(m) | \mathcal{S}_i(m)) &= \sum_{l=1}^L \rho_{il} \mathcal{N}(\mathbf{x}(m); \mu_{\mathbf{x}, il}, \mathbf{V}_{\mathbf{x}\mathbf{x}, il}), \\ \mathbf{V}_{\mathbf{x}\mathbf{x}, il} &= (\mathbf{A}_{\mathbf{y}\mathbf{y}, il} - \mathbf{A}_{\mathbf{y}\mathbf{x}, il} \mathbf{A}_{\mathbf{x}\mathbf{x}, il}^{-1} \mathbf{A}_{\mathbf{x}\mathbf{y}, il})^{-1}. \end{aligned}$$

The a priori probability $p(\mathcal{S}_i(m), \mathbf{X}(m-1)) = \alpha_i(m)$ is determined recursively according to the equation [7, 8]

$$\begin{aligned} \alpha_i(1) &= P(\mathcal{S}_i), \\ \alpha_i(m+1) &= \sum_{j=1}^{N_S} \alpha_j(m) p(\mathbf{x}(m) | \mathcal{S}_j(m)) \\ &\cdot P(\mathcal{S}_i(m+1) | \mathcal{S}_j(m)). \end{aligned} \quad (12)$$

The estimation rule (11) can be interpreted as a *cascaded estimation* rule: first the state-dependent expectation of \mathbf{y} is calculated for each state, followed by an individual weighting with the respective a posteriori probabilities $P(\mathcal{S}_i(m) | \mathbf{X}(m))$.

4.1. References: Classification Rules

The new estimation rule (11) has been compared to several HMM-based (hard resp. soft) classification rules:

$$\mathbf{ML}: \hat{\mathbf{y}} = \hat{\mathbf{y}}_{i_{\text{ML}}} \text{ with } i_{\text{ML}} = \arg \max_{i=1}^{N_S} p(\mathbf{x}(m) | \mathcal{S}_i(m)),$$

$$\mathbf{MAP}: \hat{\mathbf{y}} = \hat{\mathbf{y}}_{i_{\text{MAP}}} \text{ with } i_{\text{MAP}} = \arg \max_{i=1}^{N_S} P(\mathcal{S}_i(m) | \mathbf{X}(m)),$$

$$\mathbf{MMSE}: \hat{\mathbf{y}} = \sum_{i=1}^{N_S} \hat{\mathbf{y}}_i P(\mathcal{S}_i(m) | \mathbf{X}(m)).$$

All of these classification rules build on the fact that $\hat{\mathbf{y}}_i = E\{\mathbf{y} | \mathcal{S}_i\}$ due to the definition of the HMM states. For a detailed description of these rules please refer to [19, 7].

5. EVALUATION

The new estimation method has been evaluated both by instrumental performance measures and by informal listening tests. The complexity of the HMM was varied between $N_S = 2 \dots 64$ states, and the number of Gaussians per state was specified to be in the range $L = 1 \dots 16$. The 15-dimensional feature vector \mathbf{x} consisted of the first ten normalized auto-correlation coefficients, the zero crossing rate, the normalized frame energy, the gradient index, the local kurtosis, and the narrowband spectral centroid. Please refer to [7, 8] for a detailed description of these features.

The measurements were performed using the SI100 corpus of the BAS (*Bavarian Archive for Speech Signals*) database containing continuous German speech from 101 male and female speakers. For training about 70% of the utterances were separated, resulting in more than 25 hours of training data. The remaining speech samples constituted the test set. The speaker-dependent models were trained individually using only training data from a single speaker.

The instrumental performance evaluation was performed in terms of the *root mean square (RMS) log spectral distortion (LSD)* \bar{d}_{LSD} of the estimated spectral envelope within the extension band. As described in [4, 7] this sub-band spectral distortion measure can be determined by calculating the mean square estimation error

$$\bar{d}_{LSD} = \frac{\sqrt{2} 10}{\log_e 10} \sqrt{E\{||\mathbf{y}_{eb} - \hat{\mathbf{y}}_{eb}||^2\}}. \quad (13)$$

Note that the LSD measure is applied only to the high frequency extension band in our case.

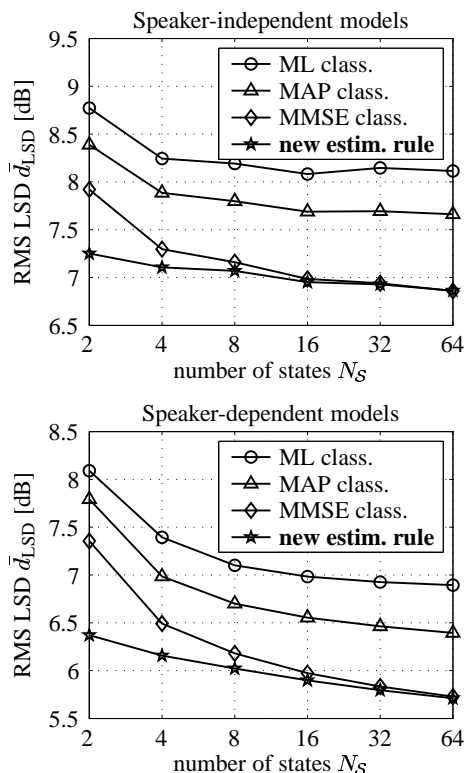


Fig. 2. Mean log spectral distortion of the estimated spectral envelope within the extension frequency band.

The results are illustrated in Fig. 2. For both speaker-independent and speaker-dependent modeling, the new MMSE estimation rule from Eq. (11) yields a consistent improvement in comparison to the classification methods. Listening tests confirmed that the new method produces an improved speech quality with less artifacts.

6. CONCLUSION

A new HMM-based algorithm for the estimation of the gain and the shape of the spectral envelope within the extension band was

proposed. The new estimation rule takes into account the a posteriori state probabilities as well as state-specific dependencies on the narrowband features. By both instrumental measurements and informal listening tests it was found that the new approach delivers a consistent improvement as compared to previously published HMM-based classification methods.

7. REFERENCES

- [1] S. Voran, "Listener ratings of speech passbands," in *IEEE Speech Coding Worksh.*, Pocono Manor, PA, USA, Sept. 1997, pp. 81–82.
- [2] 3GPP TS 26.171, "AMR wideband speech codec; general description," Mar. 2001.
- [3] 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*, Orlando, FL, USA, May 2002, vol. 1, pp. 525–528.
- [4] 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, vol. 1, pp. 237–240.
- [5] H. Carl, *Untersuchung verschiedener Methoden der Sprachkodierung und eine Anwendung zur Bandbreitenvergrößerung von Schmalband-Sprachsignalen*. Ph.D. thesis, Ruhr-Universität Bochum, Bochum, Germany, 1994, (in German).
- [6] 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.
- [7] P. Jax, *Enhancement of Bandlimited Speech Signals: Algorithms and Theoretical Bounds*. Ph.D. thesis, RWTH Aachen. vol. 15 of P. Vary (ed.), Aachener Beiträge zu digitalen Nachrichtensystemen, 2002.
- [8] P. Jax and P. Vary, "On artificial bandwidth extension of speech signals," accepted for publication in *Signal Processing (EURASIP)*, 2003.
- [9] J. Epps and W. H. Holmes, "A new technique for wideband enhancement of coded narrowband speech," in *IEEE Speech Coding Worksh.*, Porvoo, Finland, Sept. 1999, pp. 174–176.
- [10] N. Enbom and W. B. Kleijn, "Bandwidth expansion of speech based on vector quantization of the mel frequency cepstral coefficients," in *IEEE Speech Coding Worksh.*, Porvoo, Finland, Sept. 1999, pp. 171–173.
- [11] C. Avendano, H. Hermansky, and E. A. Wan, "Beyond Nyquist: Towards the recovery of broad-bandwidth speech from narrow-bandwidth speech," in *Proc. of EUROSPEECH*, Madrid, Spain, Sept. 1995, vol. 1, pp. 165–168.
- [12] S. Chenoukh, A. Gerrits, G. Miet, and R. Sluijter, "Speech enhancement via frequency bandwidth extension using line spectral frequencies," in *Proc. of ICASSP*, Salt Lake City, UT, USA, May 2001, vol. 1, pp. 665–668.
- [13] K.-Y. Park and H. S. Kim, "Narrowband to wideband conversion of speech using GMM-based transformation," in *Proc. of ICASSP*, Istanbul, Turkey, June 2000, vol. 3, pp. 1847–1850.
- [14] P. Jax and P. Vary, "Wideband extension of telephone speech using a hidden Markov model," in *IEEE Speech Coding Worksh.*, Delavan, WI, USA, Sept. 2000, pp. 133–135.
- [15] U. Kornagel, "Spectral widening of the excitation signal for telephone-band speech enhancement," in *Proc. of IWAENC*, Darmstadt, Germany, Sept. 2001, pp. 215–218.
- [16] J. A. Fuemmeler and R. C. Hardie, "Techniques for the regeneration of wideband speech from narrowband speech," in *IEEE Worksh. Nonlin. Signal and Image Proc.*, Baltimore, MD, USA, June 2001.
- [17] S. V. Vaseghi, *Advanced Signal Processing and Digital Noise Reduction*, Wiley, Teubner, 1996.
- [18] S. Kotz, N. Balakrishnan, and N. L. Johnson, *Continuous Multivariate Distributions, Volume 1: Models and Applications*, John Wiley and Sons, Berlin, Heidelberg, 2000.
- [19] P. Jax and P. Vary, "Enhancement of bandlimited speech signals," in *Proc. of Intl. ITG Conf. on Source and Channel Coding*, Berlin, Germany, Jan. 2002, pp. 347–352.