# Codebook-based Bandwidth Extension with Side-Information

Jürgen Freudenberger, Michael Müller
*Institut für Systemdynamik, HTWG Konstanz, Email: jfreuden@htwg-konstanz.de, Web: www.isd.htwg-konstanz.de*

## Introduction

With the Bandwidth Extension (BWE) techniques it is possible to enhance the bandwidth of limited speech signals and provide backward compatibility with respect to the communication networks. Artificial Bandwidth Extension (ABWE) uses correlations between the low and the high frequency ranges, to predict a wideband signal outgoing from a narrowband signal [1, 2]. The advantage of the ABWE technique is that only the receiver has to be adapted. However, ABWE only provides good quality if the correlations between the low and high frequencies are large. Especially in noisy scenarios like hands-free car communication the correlation decreases and with it the quality of the BWE. Therefore, the speech quality of ABWE is clearly affected in comparison with a wideband speech transmission [1].

This work presents a speech transmission technique using bandwidth extension with side information (BWES). To ensure backwards compatibility, this side information is embedded into the narrowband speech signal. Therefore, methods for audio steganography could be used. [3] introduces a spread spectrum approach for $\mu$-law companded speech signals. A generic approach for embedding the side information in the narrowband signal could be found in [4]. In [5], Ding presented a simple approach for the *A-law* companding. A-law companding is standardized in the ITU-T G.711 standard and used in telecommunication systems like ISDN networks or the Bluetooth speech link. The approach in [5] requires 24 bits of side information for each speech frame. This side information is embedded in the A-law encoded speech data using simple least significant bit (LSB) steganography. The LSB steganography constitutes an adequate method to embed the side information in the narrowband signal. In this work we present an approach to reduce the number of bits for the side information. The goal of the presented work is therefore to provide a best possible speech quality with as little side information as possible.

## Architecture

The BWES system consists of an encoder in the transmitter terminal and a decoder in the receiver terminal. The encoder, depicted in Figure 1, is responsible for the calculation and embedding of the speech features. To achieve a minimum number of bits, the encoder analyses the quality of an artificial BWE without additional speech features. If the ABWE receives an adequate quality no additional side information has to be used. Otherwise, additional speech features for the BWE have to be estimated.

Therefore, the recorded wideband speech signal $x_{in}$ will be split up into the high frequency signal $x_{high}$ and the low frequency signal $x_{low}$. In the block *signal extraction*, the excitation signal $e_{nb}$ and the spectral envelop $a_{nb}$ are estimated from the band limited signal $x_{low}$. In the *estimation quality verification* block the spectral envelope of the upper frequency range $\hat{a}_{hb}$ is estimated from the lower frequency band with a codebook approach. With the distance between the estimated envelope $\hat{a}_{hb}$ and the original envelope $a_{hb}$, calculated with the Mahalanobis distance, the quality of the enhancement of the spectral envelope could be estimated. With insufficient quality of the ABWE the optimal codebook entry will be transmitted. Therefore, the best fitting codebook entry is determined based on $x_{high}$.

The embedding of the side information will be done in the block *steganography*. In the A-law encoded NB signal $x_{low}$ the side information will be embedded with the LSB steganography. Therefore the SNR of $x_{low}$ decreases depending on the number of embedded bits but the NB signal is compatible with the number of conventional end-user equipment.

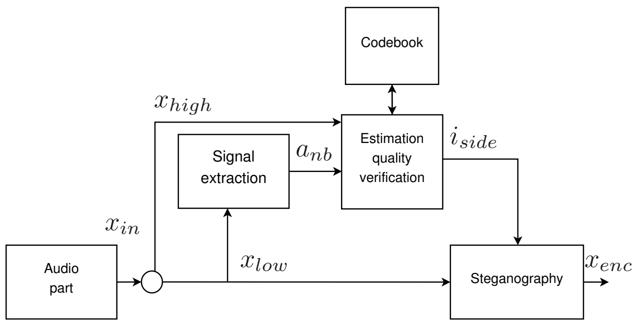


**Fig. 1:** Encoder

The decoder is illustrated in Figure 2. On the decoder side the embedded side information will be extracted from the transmitted A-law coded signal $x_{enc}$. After the signal extraction the spectral envelope and the excitation signal will be enhanced separately and the high frequencies are estimated. To improve the BWE the transmitted side information $i_{sideinfo}$ is used. Joined with the transmitted low frequencies the bandwidth extended signal $x_{recon}$ is obtained. In contrast to the estimation of the excitation signal the broadband spectral envelope cannot always be estimated satisfyingly out of the narrowband signal. In this case, the estimation can be enhanced with embedded side information. To provide this side information, we use a codebook based approach with vector quantization of the feature vectors. This approach determines a codebook which is trained on the wideband spectral envelope. To generate a representative codebook with sufficient trained centroids we used a

| | Scenario 1 | | | Scenario 2 | | |
|---|---|---|---|---|---|---|
| | S-MOS | N-MOS | G-MOS | S-MOS | N-MOS | G-MOS |
| ABWE | 3,38 | 1,64 | 2,16 | 2,37 | 3,35 | 2,37 |
| BWES | 4,19 | 1,66 | 2,71 | 3,11 | 3,39 | 2,82 |
| AMR-WB | 4,34 | 1,63 | 2,8 | 3,12 | 3,48 | 3,12 |

**Tab. 1:** BWE in noise scenarios (speech (S-MOS), noise (N-MOS), overall quality (G-MOS))

speech database with 2300 words from 40 male and female speakers. For the training of the codebook the LBG algorithm was used [6]. We used cepstrum coefficients for the representation of the spectral envelope as defined in [2].

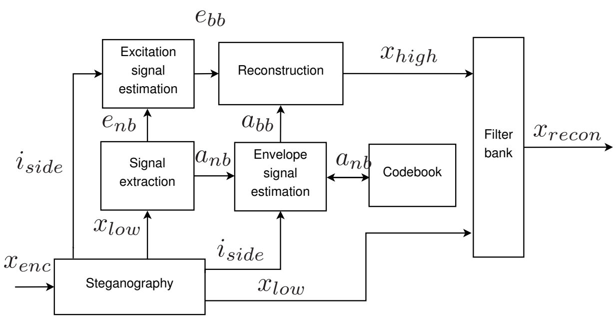


**Fig. 2:** Decoder

## Results

In the following we compare the presented approach with artificial bandwidth extension, where both systems are based on the same codebook with $n = 128$. The side information is transmitted every 20 ms where we transmit the index of the best codebook entry and 6 bits for the power of the high band signal. The speech quality has been evaluated by means of instrumental quality measures as well as informal listening tests with 20 persons. The instrumental quality analysis of the generated audio data was done with 3QUEST [7]. For the evaluation scale we use the Mean Opinion Score (MOS). The listening tests confirmed the results for the overall speech quality presented in Tab. 1.

To evaluate the robustness of the BWE against noise and echo we simulated a hands-free car communication. The clean speech signal is distorted by car noise and reverberation. To improve the SNR, noise reduction systems could be used. The output signal of the hands-free unit is transmitted to the cellular phone over a Bluetooth link. This distorted speech signal will be bandwidth extended with the BWES. To compare BWE and WB speech transmission we use the AMR-WB codec.

For this evaluation we used two scenarios. Both scenarios consider driving noise with 140km/h. Scenario 1 does not use noise reduction and scenario 2 uses a multichannel noise reduction system. The results from Tab. 1 show that ABWE is strongly affected in scenario 1. The reason is that the ABWE shifts the low frequency distortion to the upper frequency band. Therefore ABWE is not suitable for noise scenarios without noise reduction. In both cases the BWES could improve the speech quality compared to ABWE.

## Conclusions

In this work we have presented a speech transmission technique using bandwidth extension with side information for A-law encoded speech signals. The codebook based approach requires at most 12 bits of side information per 20 ms speech frame. Embedding a small number of bits as side information in the LSB results in almost inaudible distortions. Hence, it ensures backwards compatibility for ordinary receivers without BWE decoder. The presented approach improves the speech quality for noise free and noisy speech signals compared to a narrowband transmission and to artificial bandwidth extension.

## References

[1] Peter Jax, *Enhancement of Bandlimited Speech Signals: Algorithms and Theoretical Bounds*, Ph.D. thesis, Aachen, 2002.

[2] Bernd Iser, G. Schmidt, and Wolfgang Minker, *Bandwidth Extension of Speech Signals*, vol. 13 of *Lecture Notes in Electrical Engineering*, Springer, Boston (USA), 2008.

[3] Libo Zhang, Heping Ding, and S. Krishnan, "Data embedding in mu-law speech with spread spectrum techniques," in *Global Telecommunications Conference, 2005. GLOBECOM '05. IEEE*, 2005, vol. 4, pp. 4 pp. –2163.

[4] Bernd Geiser, Peter Jax, and Peter Vary, "Artificial Bandwidth Extension of Speech Supported by Watermark-Transmitted Side Information," *Integration The VLSI Journal*, vol. 10, no. 2, pp. 1497–1500, 2005.

[5] Heping Ding, "Wideband audio over narrowband low-resolution media," *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. I–489–92.

[6] Y. Linde, A. Buzo, and R. Gray, "An Algorithm for Vector Quantizer Design," *IEEE Transactions on Communications*, vol. 28, no. 1, pp. 84–95, Jan. 1980.

[7] H.W. Gierlich, F. Kettler, S. Poschen, and J. Reimes, "A new objective model for wide- and narrowband speech quality prediction in communications including background noise," in *16th European Signal Processing Conference (EUSIPCO 2008)*, 2008.