International Workshop on Acoustic Echo and Noise Control (IWAENC2003), Sept. 2003, Kyoto, Japan

IMPROVED ARTIFICIAL LOW-PASS EXTENSION OF TELEPHONE SPEECH

Ulrich Kornagel

Signal Theory, Institute for Communication Technology
Darmstadt University of Technology
Merckstrasse 25
D-64283 Darmstadt, Germany
e-mail: ulrich.kornagel@nt.tu-darmstadt.de

ABSTRACT

In this contribution a new method for artificial bandwidth extension of telephone speech towards frequencies below 300 Hz is proposed. The method defines a self-contained module which can be applied to any artificial bandwidth extension system, designed for extending the bandwidth only to frequencies *above* the telephone-band. At this, the amplification of a separate ARMA filter with a pass-band characteristic below the telephone-band is adapted such that the spectral tilt is continued to the frequencies below the telephone-band. In comparison with traditional artificial extension systems based on codebook approaches the new method produces less roughness and rattling background noise in the synthesized signal.

1. INTRODUCTION

In most telephone systems the transmitted speech is reduced to a bandwidth from 0.3 kHz to 3.4 kHz which leads to the well known speech quality degradation. To enhance the quality without transmitting extra information an algorithm for artificial bandwidth extension can be applied at the receiving end. The goal of such an algorithm is the generation of spectral components within the frequency ranges below 0.3 kHz (the low-pass extension) and from 3.4 kHz typically up to 8 kHz (the high-pass extension), i.e. complementary to the telephone-band. In recent years different approaches have been investigated (e.g. [2, 5, 6, 8, 9]) which focus on the extension towards the higher frequencies. Other publications (e.g. [3, 4, 10]) also consider the extension towards the lower frequencies. Especially in [3] the low-pass extension on the basis of a codebook approach is discussed. The speech disturbances occurring at this approach can be confirmed by the author of this paper and are the motivation for the new low-pass extension idea presented here. This new method can be linked to any traditional bandwidth extension system designed for high-pass extension only.

2. TRADITIONAL EXTENSION SYSTEMS

The theoretical background of most artificial bandwidth extension systems is a simplified version of the linear model of speech production which is closely connected with the LPC analysis. The latter divides a signal into the spectral envelope and the spectrally flat residual signal. The processing is done frame by frame with a frame length of 20 ms and a frame overlap of 10 ms for example. Both analysis components are defined for the telephone-band and have to be extended to the new bandwidth separately as described in Section 2.1 and Section 2.2. According to the linear model of speech production, the extended spectral envelope defines the wide-band shaping filter which is driven by the extended residual signal. Finally, the synthesized speech signal and the original telephone-band signal are added to build the desired wide-band signal, see Fig. 1.

2.1. Extension of the residual signal

The extended residual signal which is used as the excitation signal has to be spectrally flat according to the linear model of speech production. Furthermore, for voiced sounds the harmonic pitch structure should be maintained as well as possible. The residual signal extension towards the frequencies below 300 Hz can be done by applying a nonlinearity. One example is the quadratic function which is easy to handle as described in [7]. The upper frequency components in the range from 3.4 kHz up to 8 kHz can be obtained by a modulation approach, e.g. [7, 8].

2.2. Extension of the spectral envelope

There are several approaches that accomplish this task. A representative method which is based on a classification approach applies a pair of codebooks containing a limited number of sounds. The first codebook stores several telephone-band features like the spectral envelope, the short-time energy and the zero crossing rate for each sound. The second codebook represents the same sounds but stores the

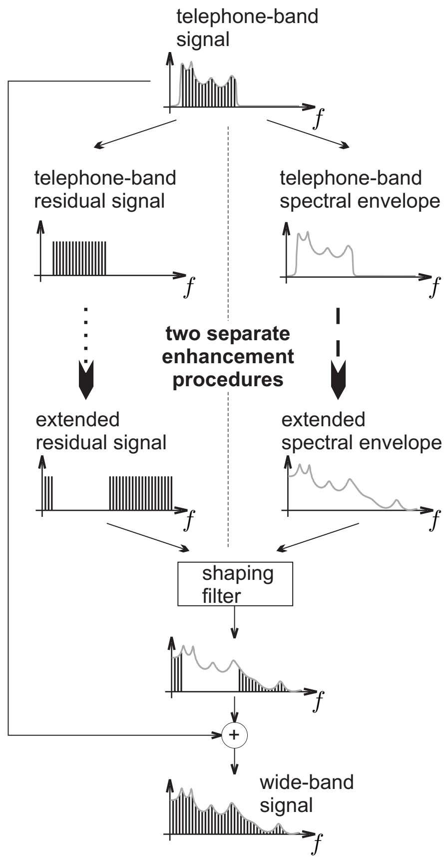

Fig. 1. *Principle of artificial wide-band extension.*

respective wide-band envelopes which define the shaping filter characteristic above 3.4 kHz (for high-pass extension) and below 0.3 kHz (for low-pass extension). In the simplest version the current features of the input telephone-band signal are extracted and compared with each telephone-band codebook entry. The classification algorithm selects the number of the codebook entry whose telephone-band features match best the current input features. Finally, with the knowledge of the entry number the desired wide-band envelope can be taken from the second codebook to define the shaping filter. This simple method, e.g. applied in [3, 9], is the basis of several more sophisticated versions, e.g. [6, 8, 10]. The drawback of this procedure is described in Section 2.3.

2.3. Drawback

The algorithm for artificial bandwidth extension is restricted to the telephone-band information to deduce on the wide-band signal. Usually, there is enough information for a reliable classification. However, several codebook entries which are quite similar in their telephone-band features differ considerably concerning the wide-band spectral envelope. Since these codebook entries are difficult to distinguish, they are occasionally selected alternately for consecutive frames in the speech production process due to classification inaccuracies. This jumpy behaviour leads to audible disturbances in the synthesized low-pass and high-pass speech signals. The synthetic low-pass signal reveals a rough and rattling background noise which also was observed in [3]. Although the speech quality can be considerably improved by more sophisticated classification methods there is still a demand of improvement especially for the low-pass extension.

3. LOW-PASS EXTENSION MODULE

To reduce the disturbances described in Section 2.3, a separate ARMA filter with a fixed pass-band characteristic below 300 Hz is applied. The only degree of freedom is an additional amplification adjustment factor ξ which is calculated such that the spectral tilt is continued to the frequencies below 300 Hz by means of the adjusted ARMA filter. (The basic idea of continuing the spectral tilt is also shown in [1].) The spectral tilt is extracted from the wide-band envelope obtained by a traditional bandwidth extension system with high-pass extension only, see Fig. 2. The desired

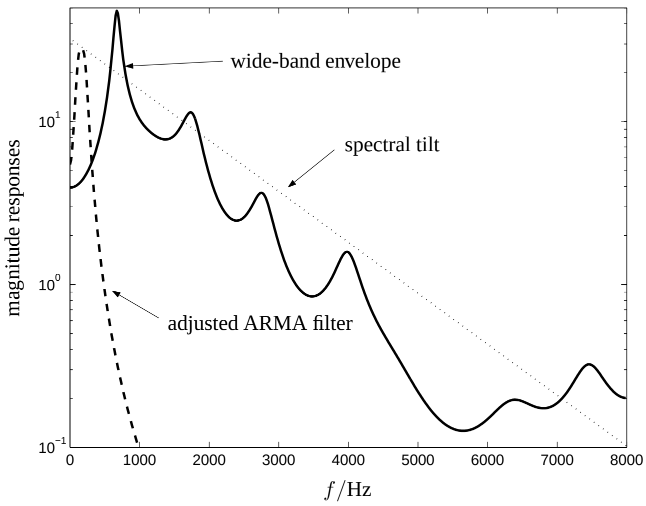

Fig. 2. *Continuation of spectral tilt towards the frequencies below* 300 Hz *by means of the adjusted ARMA filter.*

low-pass signal is generated by driving the adjusted ARMA filter with the low-pass excitation signal (see Section 2.1 and Section 3.1).

3.1. Power of low-pass excitation signal

The high-pass extended wide-band signal obtained by the traditional extension system can be divided into the wide-band envelope (see Fig. 2) and the wide-band residual signal. The latter is spectrally flat and has spectral components above 0.3 kHz (up to 8 kHz). The low-pass excitation signal has to be amplified such that its power divided by 0.3 kHz equals the power of the wide-band residual signal divided by 7.7 kHz (8 kHz - 0.3 kHz). (This amplification is independent of the ARMA filter adjustment which only considers the wide-band envelope shown in Fig. 2.)

3.2. ARMA filter

The fixed filter is defined to be an ARMA filter $G_{ARMA}(z)$ with three zeros at $z = 0$, one zero at $z = 0.98$, two complex conjugate poles at $z = 0.98e^{\pm j(2\pi 120Hz)/f_s}$, and two complex conjugate poles at $z = 0.98e^{\pm j(2\pi 215Hz)/f_s}$ with $f_s = 16$ kHz (cp. Eq. 1), since the mean pitch frequency for male voices is 120 Hz and for female voices is 215 Hz.

$$G_{ARMA}(z) = \frac{1.0 - 0.98z^{-1}}{1.0 - 3.9108z^{-1} + 5.7445z^{-2} - 3.7560z^{-3} + 0.9224z^{-4}} \quad (1)$$

The lower frequency part of the magnitude response $\left|G_{ARMA}\left(e^{j\Omega}\right)\right|$ is depicted in Fig. 3. The attenuation

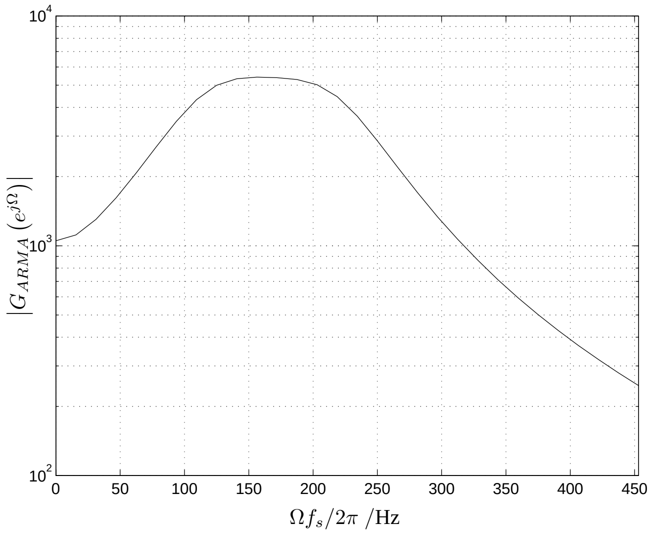

Fig. 3. *Lower frequency range of the magnitude response* $\left|G_{ARMA}\left(e^{j\Omega}\right)\right|$.

above 300 Hz is not a crucial item since the low-pass excitation signal is already restricted to the frequency range below 300 Hz.

3.3. Adaptation of amplification

The factor ξ for the adjusted ARMA filter

$$G_{ad}\left(e^{j\Omega}\right) = \xi \cdot G_{ARMA}\left(e^{j\Omega}\right)$$

is calculated separately for each frame. At this, the starting point is the wide-band envelope (see Section 3) which is defined by a 16^{th} order all-pole filter $H_{16}(z)$ (cp. Eq. 2), as an example.

$$H_{16}(z) = \frac{1.0}{1.0 + \sum_{i=1}^{16} a_i z^{-i}} \quad (2)$$

The adjustment factor ξ is defined only by the spectral tilt which is extracted from $H_{16}(z)$. This can be done by reducing the order of $H_{16}(z)$ to get the *first* order all-pole filter $H_1(z)$, cp. Eq. 3.

$$H_1(z) = \frac{\tilde{b}_0}{1.0 + \tilde{a}_1 z^{-1}} \quad (3)$$

Since $H_1(z)$ has a real coefficient $\tilde{a}_1$ the pole occurs on the real axis in the z-domain. For voiced sounds, the spectral tilt is declining which forces the pole to occur somewhere on the positive real axis. Unvoiced sounds show a flatter or a rising spectral tilt towards the higher frequencies and consequently the pole moves towards the negative real axis. As an example, Fig. 4 and Fig. 5 show the magnitude responses of $H_{16}\left(e^{j\Omega}\right)$ and $H_1\left(e^{j\Omega}\right)$ for a voiced sound ($H_{16}^{(v)}\left(e^{j\Omega}\right)$ and $H_1^{(v)}\left(e^{j\Omega}\right)$) and an unvoiced sound ($H_{16}^{(uv)}\left(e^{j\Omega}\right)$ and $H_1^{(uv)}\left(e^{j\Omega}\right)$), respectively.

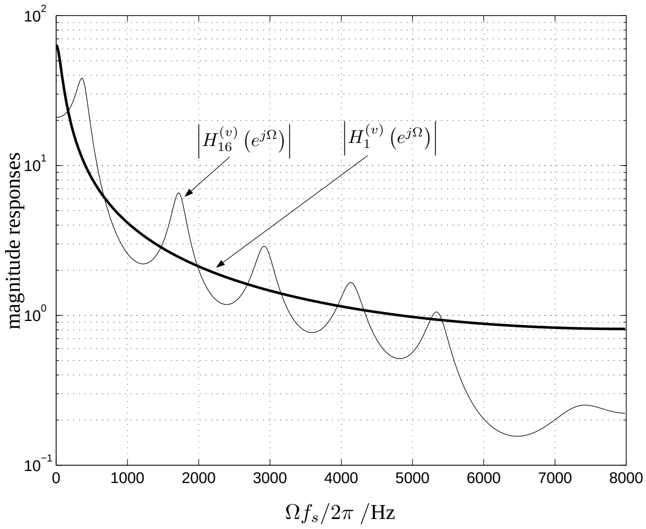

Fig. 4. *Magnitude responses* $\left|H_{16}^{(v)}\left(e^{j\Omega}\right)\right|$ *and* $\left|H_1^{(v)}\left(e^{j\Omega}\right)\right|$ *for a voiced sound.*

Both coefficients $\tilde{a}_1$ and $\tilde{b}_0$ of $H_1(z)$ can be obtained by applying the step-down recursion which maps the 16 filter coefficients ($a_1 \ldots a_{16}$) of $H_{16}(z)$ to the 16 reflection coefficients ($\Gamma_1 \ldots \Gamma_{16}$). The coefficients of $H_1(z)$ are then defined according to Eqs. 4 and 5.

$$\tilde{a}_1 = \Gamma_1 \quad (4)$$

$$\tilde{b}_0 = \frac{1.0}{\sqrt{\prod_{i=2}^{16}\left(1.0 - |\Gamma_i|^2\right)}} \quad (5)$$

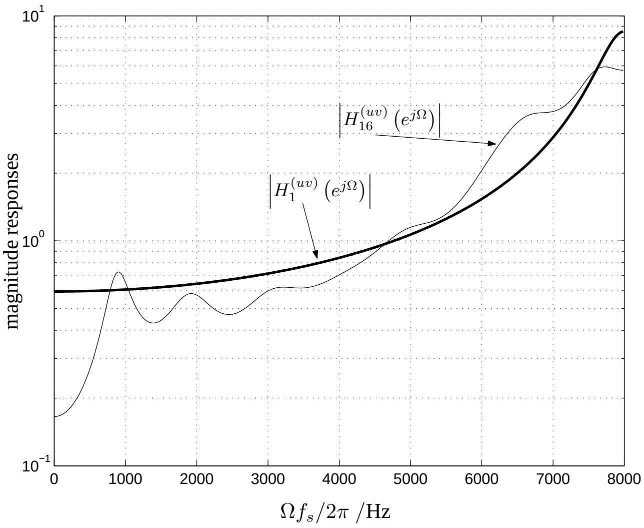

Fig. 5. *Magnitude responses* $\left|H_{16}^{(uv)}\left(e^{j\Omega}\right)\right|$ *and* $\left|H_{1}^{(uv)}\left(e^{j\Omega}\right)\right|$ *for an unvoiced sound.*

The adjustment factor ξ is finally calculated as the ratio of $\left|H_1\left(e^{j\Omega_c}\right)\right|$ and $\left|G_{ARMA}\left(e^{j\Omega_c}\right)\right|$ at one specific frequency Ω_c below 300 Hz, e.g. at $\Omega_c = 2\pi\,150$ Hz$/f_s$, see Eq. 6.

$$\xi = \frac{\left|H_1\left(e^{j\Omega_c}\right)\right|}{\left|G_{ARMA}\left(e^{j\Omega_c}\right)\right|} \tag{6}$$

3.4. Postprocessing

Due to estimation inaccuracies the factor ξ sometimes reveals noisy behaviour over the time which again would lead to a roughness in the synthesized signal. This can be easily remedied by smoothing the factor over the time. Since the influence of ξ on the extension process is obvious, further, more sophisticated postprocessing ideas can be easily applied.

4. RESULT AND CONCLUSION

The proposed method allows the spectral extension towards frequencies below 300 Hz. At this, an ARMA filter with a pass-band characteristic below 300 Hz is adapted such that its magnitude and the magnitude defined by the spectral tilt are equal at one specific frequency below 300 Hz. In comparison with a codebook approach, the proposed method produces speech with less rough and rattling background noise at the lower frequencies. Furthermore, the influence of the adapted ARMA filter amplification on the synthesized speech is straightforward and thus postprocessing is a simple task. Since the proposed method is self-contained it can be linked to any traditional bandwidth extension system designed for high-pass extension only.

5. REFERENCES

[1] H. Yasukawa: *Restoration of Wide Band Signal From Telephone Speech Using Linear Prediction Error Processing*, Proc. International Conference on Spoken Language Processing, Philadelphia, PA, USA, Vol. II, pp. 901 - 904, October 1996.

[2] N. Enbom, W. B. Kleijn: *Bandwidth Expansion of Speech Based on Vector Quantization of the MEL Frequency Cepstral Coefficients*, Proc. IEEE Workshop on Speech Coding 1999, Porvoo, Finland, pp. 171 - 173, June 1999.

[3] G. Miet, A. Gerrits, J. C. Valière: *Low-Band Extension of Telephone-Band Speech*, Proc. ICASSP 2000, Istanbul, Turkey, Vol. III, pp. 1851 - 1854, June 2000.

[4] K.-Y. Park, H. S. Kim: *Narrowband to Wideband Conversion of Speech Using GMM Based Transformation*, Proc. ICASSP 2000, Istanbul, Turkey, Vol. III, pp. 1843 - 1846, June 2000.

[5] M. Mason, D. Butler, S. Sridharan, V. Chandran: *Narrowband Speech Enhancement Using Fricative Spreading and Bandwidth Extension*, Proc. ISPACS 2000, Honolulu, Hawaii, USA, pp. 714 - 717, Nov. 2000.

[6] S. Chennoukh, A. Gerrits, G. Miet, R. Sluijter: *Speech Enhancement via Frequency Bandwidth Extension Using Line Spectral Frequencies*, Proc. ICASSP 2001, Salt Lake City, Utah, USA, Vol. I, pp. 665 - 668, May 2001.

[7] U. Kornagel: *Spectral Widening of the Excitation Signal for Telephone-Band Speech Enhancement*, Proc. IWAENC 2001, Darmstadt, Germany, pp. 215 - 218, September 2001.

[8] P. Jax, P. Vary: *Enhancement of Band-Limited Speech Signals*, 10th Aachen Symposium on Signal Theory, Aachen, Germany, pp. 331 - 336, September 2001.

[9] J. A. Fuemmeler, R. C. Hardie, W. R. Gardner: *Techniques for the Regeneration of Wideband Speech from Narrowband Speech*, EURASIP Journal on Applied Signal Processing, Vol. 2001, Nr. 4, pp. 266 - 274, December 2001.

[10] U. Kornagel: *Spectral Widening of Telephone Speech Using an Extended Classification Approach*, Proc. EUSIPCO 2002, Toulouse, France, Vol. II, pp. 339 - 342, September 2002.