SPECTRAL WIDENING OF THE EXCITATION SIGNAL FOR TELEPHONE-BAND SPEECH ENHANCEMENT

Ulrich Kornagel

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

ABSTRACT

In current telephone systems the speech to be transmitted is band-limited. The resulting speech quality degradation can be reduced by supplementing the missing spectral components of an enlarged frequency band in an artificial way. One important component of an enhancement system is the wide-band synthesis-filter that has to be driven with a wide-band excitation signal. The focus of this paper is to propose different methods to generate this wide-band excitation signal from a telephone-band limited version.

1. INTRODUCTION

In current telephone systems the transmitted speech is usually reduced to the frequency band from 0.3 kHz to 3.4 kHz. This results in the typical telephone sound. Since wide-band speech with a frequency range from the lower hearing threshold up to 8 kHz sounds more natural, it is desirable to enhance the bandwidth of the telephone-band signal according to this range in an artificial way. As a result one can get high quality speech at the cost of telephone speech.

The basic idea, as proposed in several publications [2, 3, 4, 5, 6], is to create a database of wide-band speech connected with a database of telephone-band speech. The missing information of the telephone-band signal to be enhanced is then taken out of the wide-band speech database controlled on the basis of the telephone-band database.

The artificial spectral components are synthesized by an AR-model (for instance), thus, the model parameters and the excitation signal are needed. The model parameters are taken from the wide-band database. However, to get the excitation signal this principle is not feasible.

One possible starting point of computing the excitation signal is to apply a linear prediction error filter to the telephone-band signal. This leads to the telephone-band residual signal that now has to be enhanced.

One very simple method to achieve the enhancement towards the higher frequencies is to upsample the telephone-band residual signal by a factor of two ([1], [2], [5], [6]), i.e., by duplicating the sampling frequency by including a zero sample after each sample (without a subsequent anti-imaging filter as needed for an interpolation). This leads to a simple spectral duplication of the residual signal. The disadvantage of this method is that the resulting excitation signal only contains spectral components in the frequency range from 0.3 kHz to 3.4 kHz and from 4.6 kHz to 7.7 kHz. Thus, the enhancement towards the upper frequencies is only partial. Furthermore, tonal disturbances can occur because the harmonic pitch structure in the frequency domain for voiced sounds is usually disturbed. This paper presents two methods to deal with these problems.

The enhancement towards the lower frequencies can be achieved by the application of a non-linearity [2]. This idea is picked up and refined in this paper.

2. LOWPASS- AND HIGHPASS-REGENERATION

The basic demand is to use as much telephone-band data as possible to ensure a natural sounding result after the later synthesis. The idea is firstly to calculate the residual signal block-by-block from the incoming telephone-band signal (at a sampling frequency of 8 kHz). This is done by means of a linear prediction error filter of low order (e.g. order ten) in a way that only the spectral envelope is flattened. Afterwards, the spectral components of the residual signal are moved to the lower and upper frequencies without leaving spectral gaps.

2.1. Hp-regeneration

At first, two ways to enhance the excitation signal towards the upper frequencies (i.e. above 3.4 kHz) are considered. This is called the hp-regeneration.

The first version, as shown in Fig. 2, focuses on the preservation of the harmonic pitch structure in the frequency domain for voiced sounds. That is, all peaks are forced to appear more or less at an integer multiple of the pitch frequency [1]. To achieve this, the residual signal is transformed by the Discrete Fourier Transform to the frequency domain where it is interpolated by a factor of two, i.e., to a sampling frequency of 16 kHz. Then the analytic signal is computed. Afterwards, the spectral components are copied to the higher frequencies, starting at 3.4 kHz, by a cyclic copying-loop. The cyclic mode means that the source frequency of the copying loop jumps back to the beginning of the telephone-band when it reaches the upper frequency of the telephone-band. This mode is necessary because the destination gap to be filled is bigger than the telephone-band. To preserve the harmonic structure, as shown in Fig. 1, the peak with the lowest frequency f_{p1} within the telephone-band has to be copied to the frequency f_{dest} of the first peak to be recovered above the telephone-band. If the frequency difference $f_{p1} - 0.3$ kHz is smaller than the difference $f_{dest} - 3.4$ kHz the reference peak within the telephone-band is not the first but the second peak at the frequency f_{p2} to avoid a spectral gap just above 3.4 kHz. This is chosen by a logic. The location of the frequencies is shown in Fig. 1. In case of unvoiced sounds

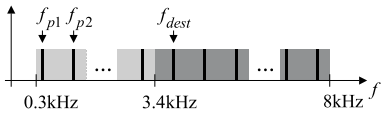

Figure 1: Location of the frequencies.

there are no peaks to care about. Then an arbitrary but reasonable pitch frequency (e.g. 50 Hz) should be chosen to assure the functioning of the algorithm. The mixed case does not need to be considered separately because the more a sound tends to be voiced the more reliable the pitch frequency can be estimated and, thus, the increased demand of preservation of the harmonic structure is satisfied.

After the copying process the spectral components in the telephone-band are discarded and then the (still) analytic signal has to be transformed back to the time domain. Finally, the computation of the real part of the signal leads to the desired excitation signal with frequency components above 3.4 kHz. An overview of this version is given in Fig. 2. On the one hand the method avoids tonal disturbances due to the disturbance of the harmonic pitch structure, on the other hand the direct manipulation in the frequency domain leads to a slight roughness in the resulting signal.

The second version of the hp-regeneration, as shown in Fig. 3, reduces this roughness. For this version, however, it is not possible to maintain the harmonic pitch structure. It uses a fixed filter of finite length to take

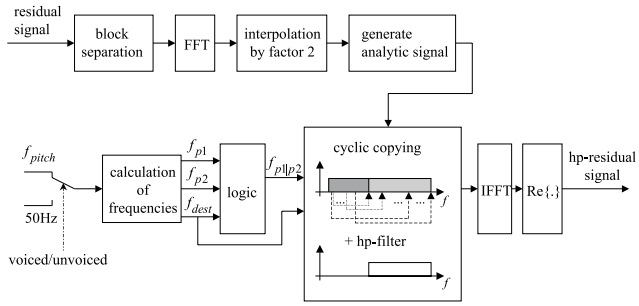

Figure 2: Hp-regeneration with pitch-control.

the spectral components out of the telephone band. These components are shifted to the upper band by two fixed frequencies.

The spectral gap to be filled is about 1.5 times larger than the telephone-band. Thus, the telephone-band should be filtered by a bandpass which has a bandwidth of half the width of the spectral gap (reduced by a small tolerance width to avoid aliasing effects due to the following shift). Therefore a bandpass with a bandwidth of 2.25 kHz is used that passes signal components in the frequency range from 0.3 kHz to 2.55 kHz. The resulting signal then has to be shifted to the upper band once by the frequency $f_1 = 3.1$ kHz and once by the frequency $f_2 = 5.35$ kHz. In practice the incoming residual signal is interpolated by a factor of two, followed by the bandpass filtering as explained above. The shifting is achieved by computing the analytic signal and modulating it with complex exponential functions of both frequencies f_1 and f_2, respectively. Finally, the real parts of both complex signals are added to get the desired excitation signal as shown in Fig. 3.

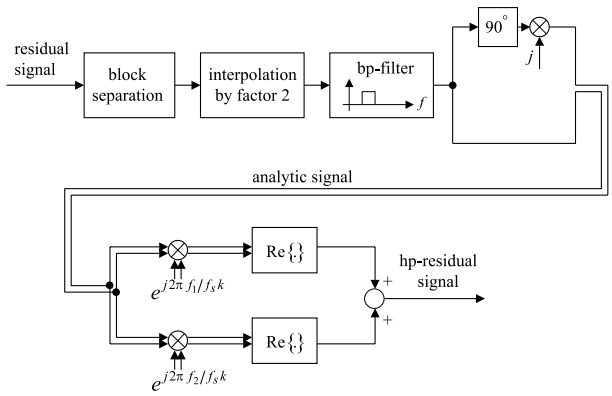

Figure 3: Hp-regeneration without pitch-control.

Both versions can be applied alternatively. However, the first version additionally requires a pitch estimator.

2.2. Lp-regeneration

Now the enhancement to the frequencies below 0.3 kHz, the so called lp-regeneration, is considered, as shown in Fig. 5. The first steps are identical to the ones of the hp-regeneration. That is, the method starts with the interpolated (by a factor of two) telephone-band residual signal. The generation of new frequency components is achieved by the application of a quadratic function [2]. It generates a harmonic structure in the frequency domain with the pitch frequency as the basic frequency. The remaining problem is how to get a nearly white (concerning the spectral envelope!) excitation signal with a power fitting to the power of the telephone-band residual signal. At this, the quadratic function in the time domain shows a useful property: The calculation can be interpreted as the inverse Fourier Transform of the convolution of the Fourier transformed signal with itself. Since the envelope of the residual signal spectrum can be modeled by a rectangular function the resulting envelope takes the shape of a triangular function within the spectral destination gap below 0.3 kHz. Additionally, the non-linearity generates frequency components above 0.3 kHz that have to be suppressed. The suppression of these components and the compensation of the triangular envelope can be achieved by filtering with a modified lowpass filter with the inverse characteristic as defined in (1).

$$|G_{lp}(f)| = \begin{cases} \frac{1}{\left(1-\frac{1}{3100}\left|f/\mathrm{Hz}\right|\right)} & \text{for } |f/\mathrm{Hz}| < 300 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

To get a well-defined time delay this ideal filter has to be approximated by a linear phase FIR filter. In this work an FIR filter of the order 200 is used as shown in Fig. 4.

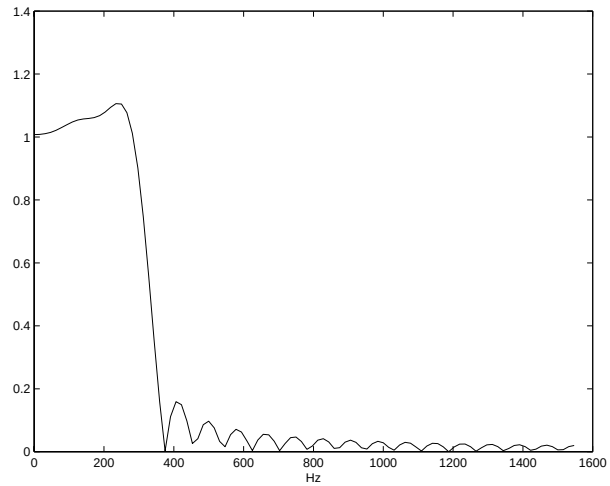

Figure 4: Modified lowpass filter.

The convolution produces an undesired additional DC-offset in the signal that easily can be calculated and removed. Finally, the power of the lowpass residual signal has to be matched to the power of the original telephone-band residual signal so that the power densities are equal. This is done by an amplification factor.

The quality of the synthesized signal is improved although a roughness is introduced due to the application of the quadratic function. The entire lp-regeneration is shown in Fig. 5.

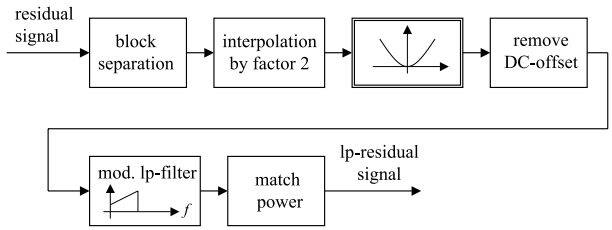

Figure 5: Lp-regeneration with quadratic function.

3. RESULTS

The following two spectrograms show the interpolated original telephone-band residual signal (Fig. 6) and the enhanced version (Fig. 7) generated by the described algorithms. The difference between the two hp-regeneration methods is hardly visible in the spectrograms, therefore, only the second method is depicted here.

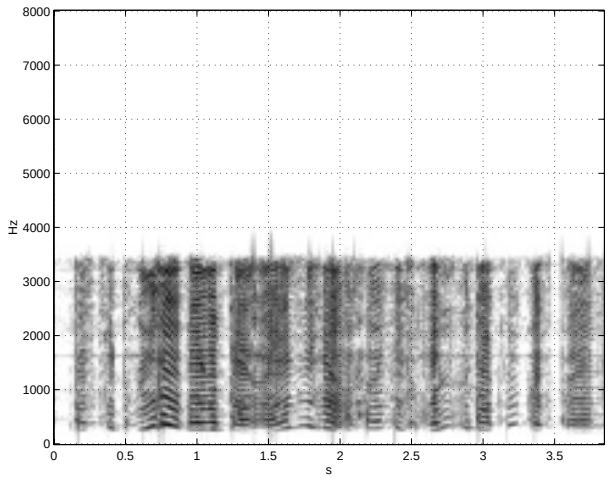

Figure 6: Original telephone-band residual signal (interpolated)

As one can see the generated wide-band residual signal has frequency components in the frequency band complementary to the telephone-band. This is because the synthesis filter (also a component of the entire speech enhancement system) has only to be fed with these frequencies. The resulting enhanced speech signal calculated as the sum of the original telephone-band signal and the artificial supplement then shows a spec-

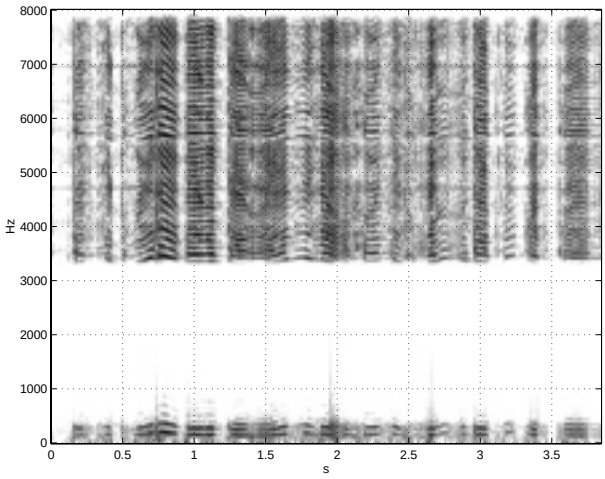

Figure 7: Enhanced residual signal

trum without gaps. In comparison with the original telephone-band speech the enhanced speech sounds more natural.

4. CONCLUSION

The principal problem addressed here was to enhance the bandwidth of a telephone-band signal to a bandwidth of about 8 kHz. For this purpose a synthesis filter taken from a wide-band speech database has to be driven with an enhanced excitation signal. Concerning the enhancement of the latter signal to the upper frequency band two methods were presented. The difference to the simple upsampling method is the possibility to maintain the harmonic pitch structure in the frequency domain in case of voiced sounds for one version and the avoidance of spectral gaps for both versions. Concerning the enhancement to the lower frequency band a refined method based on the use of a quadratic function was proposed. In comparison with the original telephone-band speech a more natural sounding speech could be achieved when the described enhancement methods were applied to the entire speech enhancement system.

5. REFERENCES

[1] Makhoul, J. and Berouti, M.: *High-frequency regeneration in speech coding systems*, IEEE Proc. ICASSP-79, pp. 428-431, Hartford, Conn., USA, 1979.

[2] Carl, H.: *Untersuchung verschiedener Methoden der Sprachcodierung und eine Anwendung zur Bandbreitenvergrößerung von Schmalband-Sprachsignalen*, Dissertation, Ruhr-Universität Bochum, 1994.

[3] Chan, C. and Hui, W.: *Quality enhancement of narrowband celp-coded speech via wideband harmonic re-synthesis*, ICASSP-97, vol. 2, pp. 1187-1190, Munich, Germany, 1997.

[4] Epps, J. and Holmes, W.H.: *A new technique for wideband enhancement of coded narrowband speech*, IEEE Workshop on Speech Coding, Proceedings, pp. 174-176, 1999.

[5] Enbom, N. and Kleijn, W.B.: *Bandwidth expansion of speech based on vector quantization of the mel frequency cepstral coefficients*, IEEE Workshop on Speech Coding, Proceedings, pp. 171-173, 1999.

[6] Jax, P. and Vary, P.: *Wideband extension of telephone speech using a hidden markov model*, IEEE Workshop on Speech Coding, Proceedings, pp. 133-135, 2000.