COMPARISON OF A DISCRETE WAVELET TRANSFORM AND A NONUNIFORM POLYPHASE FILTERBANK APPLIED FOR SPECTRAL SUBTRACTION

Andreas Engelsberg and Thomas Gülzow

Institute for Network and System Theory,
Technical Department, Kiel University,
Kaiserstrasse 2, D-24143 Kiel / Germany,

`ae@techfak.uni-kiel.de` and `tg@techfak.uni-kiel.de`

ABSTRACT

Spectral subtraction is a popular method for speech enhancement, if the speech signal is corrupted by additive noise. It is based on the manipulation of the magnitude of the noisy speech spectrum. Previous realizations used linearly spaced frequency transformations. We propose the application of two filterbanks with bark-scaled frequency bands: a discrete wavelet transform and a nonuniform polyphase filterbank. The enhancement results are compared to those obtained with uniform spectral transformations.

1. INTRODUCTION

The method of spectral subtraction is widely used for speech enhancement. It is applied to speech signals, which are disturbed by additive noise with slowly varying spectral characteristics. A practical application for example is hands-free communication in noisy environments. Spectral subtraction is performed by subtracting a mean magnitude of the noise spectrum from the disturbed spectrum to obtain an estimation of the magnitude of the noise-free spectrum.

The method may be interpreted as spectral equalizing of the noisy speech signal by applying spectral weights to the transformed signal. In practical applications the spectral analysis and synthesis is usually performed by a Discrete Fourier Transform and inverse transformation [1] with overlap-add techniques or by analysis and synthesis filterbanks, for example polyphase filterbanks [2]. The systems have a uniformly spaced division of the frequency domain.

In speech recognition systems the spectral analysis is successfully performed using spectral transformations with bark-scaled frequency bands [5]. This fact motivated the investigation of nonuniform-bandwidths filterbanks with respect to the spectral analysis of the human ear for spectral equalizing methods.

We applied two different types of filterbanks to approximate a bark-scaled frequency spacing:

1. A non-critically decimated discrete wavelet filterbank.
2. A modified polyphase filterbank with allpass transformations in the polyphase network.

The structures and functionalities of the filterbanks are described and the influence to the performance of the spectral subtraction rules due to BOLL [1] and EPHRAIM AND MALAH [4] are shown in experimental results.

2. SPECTRAL SUBTRACTION

The basic idea of spectral subtraction is applied to noisy speech signals with additive noise

$$x(k) = s(k) + n(k) \quad , \tag{1}$$

where $x(k)$ denotes the noisy signal, $s(k)$ the speech signal, and $n(k)$ the noise. The estimation of the noise-reduced speech spectrum is obtained by subtracting an estimated mean spectral magnitude $\overline{|N(e^{j\Omega})|}$ of the noise from the spectral magnitude $|X(e^{j\Omega})|$ of the noisy signal:

$$\hat{S}(e^{j\Omega}) = \left(|X(e^{j\Omega})| - \overline{|N(e^{j\Omega})|}\right) e^{j\varphi_x(\Omega)} \quad , \tag{2}$$

where $\varphi_x(\Omega)$ is the phase of the disturbed speech signal. $X(e^{j\Omega})$ and $N(e^{j\Omega})$ are the Fourier transformations of $x(k)$ and $n(k)$.

The mean magnitude of the noise spectrum is assumed to be estimated e.g. during speech pauses using a voice activity detector or by spectral-minima tracking.

As denoted in the introduction, equation (2) may be interpreted as spectral weighting of the noisy speech signal:

$$\hat{S}(e^{j\Omega}) = G(e^{j\Omega}) \cdot X(e^{j\Omega}) \quad , \tag{3}$$

where

$$G(e^{j\Omega}) = \frac{|X(e^{j\Omega})| - \overline{|N(e^{j\Omega})|}}{|X(e^{j\Omega})|} \quad . \tag{4}$$

Negative values of $G(e^{j\Omega})$ are estimation errors and therefore forced to zero.

A major drawback of spectral-subtraction systems is the remaining of residual tonal noise with unnatural sound. More sophisticated spectral weightings lead to a reduction of this phenomenon [4].

3. WAVELET TRANSFORM

As an alternative to the Fourier transform the wavelet transform can be applied for spectral analysis of a signal. The continuous wavelet transform (CWT) of a signal $x(t)$ is given by

$$\mathcal{W}_x^{\psi}(b,a) = |a|^{-\frac{1}{2}} \int_{-\infty}^{+\infty} x(t)\psi^*\left(\frac{t-b}{a}\right) dt \quad , \qquad (5)$$

where $\psi(t)$ is the prototype wavelet. By shifting and scaling $\psi(t)$ with the parameter a and b, all basis functions $\psi_{b,a}(t) = |a|^{-\frac{1}{2}}\psi\left(\frac{t-b}{a}\right)$ are obtained. Large values of a cause $\psi_{b,a}(t)$ to become a lower-frequency and dilated version of $\psi(t)$. For small a values, the function $\psi_{b,a}(t)$ becomes a contracted version of $\psi(t)$ with higher frequency components. As a consequence, the resolution in the time-frequency plane is not constant. For high frequencies the resolution of the wavelet transform is sharp in time but poor in frequency, while for small frequencies the resolution is sharp in frequency and poor in time.

In the frequency domain the wavelet transform can be interpreted as a filterbank with bandpasses whose bandwidths $\Delta\omega_i$ increase monotonously with the center frequency ω_{0_i}. It can be shown that the relative bandwidth $Q = \frac{\Delta\omega_i}{\omega_{0_i}}$ is independent on the parameter a, so the wavelet transform is called 'constant-Q' analysis. This is very similar to the frequency analysis of the human ear.

The digital realization of (5) requires the discretization of the parameters a and b. Usually they are chosen to be on a dyadic grid. Then a is a power of 2 and b is dependent on a, so that $a_m = 2^m, b_{mn} = a_m nT$, where $m, n \in \mathbb{Z}$. On this basis the wavelet transform in application to a discrete signal $x(k)$ becomes

$$w_x^{\psi}(2^m n, 2^m) = 2^{-\frac{m}{2}} \sum_k x(k)\psi^*\left(2^{-m}k - n\right) \quad , \qquad (6)$$

and realizes an octave analysis with different sampling rates in each octave.

To increase the resolution in frequency by a factor M, it is possible to use M dyadic wavelet analyses (voicing) each with the scaled prototype wavelet

$$\psi^j(k) = 2^{-\frac{j}{2M}}\psi(2^{-\frac{j}{M}}k), \quad j = 0, ..., (M-1) \quad . \qquad (7)$$

The non-critically decimated wavelet filterbank is based on the Á-Trous algorithm [6] and is one realization of (6).

In figure 1 the realized wavelet-filterbank structure with $p+1$ octaves and M voices per octave is shown. The bandpass filters $g^i(n)$, $i = 0, .., (M-1)$, are the prototype wavelets for each dyadic wavelet analysis. The decimation of 2^l in the l-th octave as shown in figure 1 allows the use of the same filter within each octave. The function of the lowpass filter $f_a(n)$ may be interpreted as an antialiasing-filter.

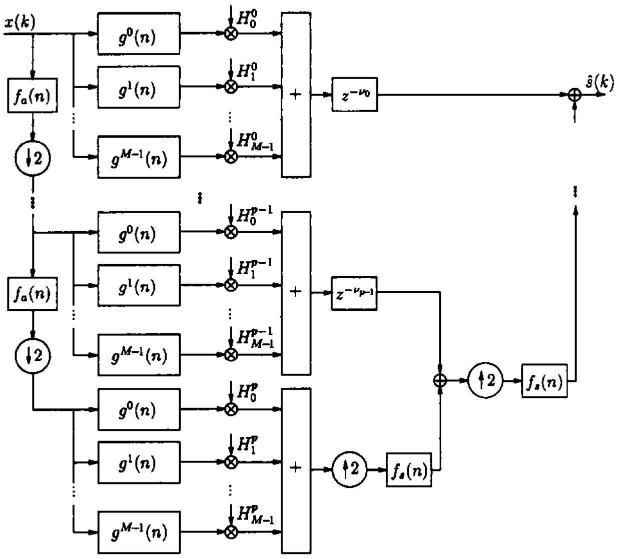

Figure 1: structure of the realized wavelet filterbank.

The synthesis filterbank interpolates the subbands to the next higher sampling rate and adds the result to the output of the next octave, taking care of the correct delay as produced in the analysis part.

4. MODIFIED POLYPHASE FILTERBANK

The analytical description of a polyphase filterbank leads to a transfer function

$$H_i(z) = \sum_{\rho=0}^{M-1} \sum_{p=0}^{\frac{L_p}{M}-1} h(pM+\rho)[z^{-1}]^{pM+\rho} e^{-ji\frac{2\pi}{M}\rho}, \qquad (8)$$

$i = 0 \ldots (M-1)$, with bandpass characteristic for each channel, where $h(k)$, $k = 0 \ldots (L_p-1)$ is the prototype lowpass. The replacement of the delays z^{-1} with an allpass of degree one

$$H_A(z) = \frac{\alpha z + 1}{z + \alpha}, \quad -1 < \alpha < 1, \qquad (9)$$

performs a transformation of the abscissa of the transfer function while the attenuation is not disturbed [3]. For $-1 < a < 0$ the bandwidths of the filters increase monotonously with the center frequency. With a suitable choice of the parameter a the frequency resolution of the human ear can be approximated. In figure 2 the structure of the allpass-transformed polyphase filterbank is shown. The long cascade of allpasses with degree one produces an output signal $\hat{s}(k)$ whose phase is disturbed. The correction of the phase requires an additional filteroperation. We choose a nonrecursiv filter whose impulse response is the limited time-inverted impulse response of the modified polyphase filterbank.

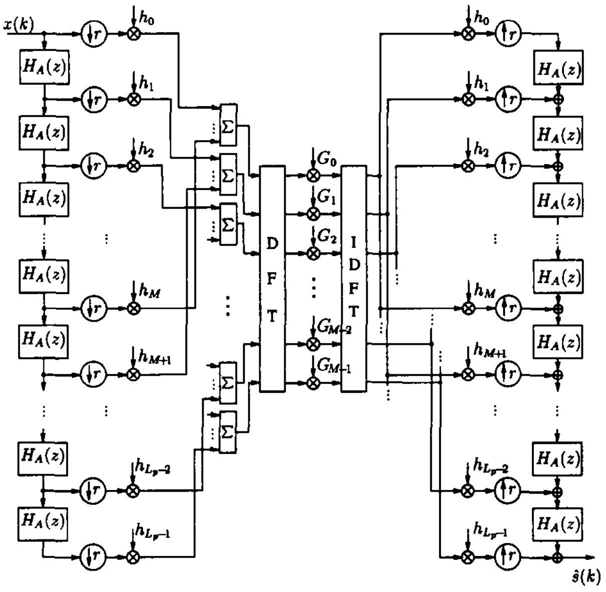

Figure 2: structure of the realized allpass-transformed polyphase filterbank.

5. EXPERIMENTS

The experiments were carried out using the above described filterbanks.

The wavelet filterbank was implemented with 7 octaves and 10 voices per octave ($p=6$ and $M=10$ in figure 1). This choice leads to a sufficient frequency resolution with 70 subbands. The prototype filter is the sampled complex Morlet wavelet

$$\psi(t) = e^{j\omega_0 t} e^{-\frac{\beta^2 t^2}{2}} \qquad (10)$$

[7] with 101 coefficients. The interpolation was implemented by FIR filters of length 71.

The modified polyphase filterbank has $M = 256$ channels. The linear-phase prototype lowpass filter was chosen to a length of $L_p = 1024$. The parameter a was set to approximate the frequency analysis of the human ear. Because of increasing bandwidths to higher frequencies the decimation factor r in figure 2 has to be limited to avoid aliasing. Listening tests showed that the factor $r = M/4$ leads to a sufficient quality of the output signal.

We examined the enhancement system with the spectral-subtraction rule due to BOLL [1] and the subtraction rule due to EPHRAIM AND MALAH [4].

In a first step we applied the nonuniform filterbanks to the basic spectral-subtraction scheme of BOLL. In this case a high amount of residual noise occurs and the repercussion of the filterbanks on the enhancement system can be well understood. The corresponding results by the application of different filterbanks are visualized by four spectrograms in figure 3. The speech signal was recorded in a running car with a sampling frequency of 11.025 kHz.

The first spectrogram shows the frequency representation of the original noisy speech signal. Most of the noise energy is located in the low-frequency area. This is characteristic for the car environment, where the noise is mostly produced by the engine.

To demonstrate the differences between uniform and nonuniform filterbanks the second spectrogram contains the frequency representation of the enhanced signal by using the allpass-transformed polyphase filterbank with $a = 0$. Thus no frequency warping has been done and the filterbank is uniform. In this case the decimation was performed with a factor 2 above critical subsampling. Especially in speech pauses the enhanced signal contains a lot of randomly distributed spectral peaks, which produce undesirable tonal residual noises.

The third spectrogram shows the result using the allpass-transformed polyphase filterbank with $a = -0.49$ and the fourth spectrogram the application of the above described wavelet filterbank as spectral transformation.

The different frequency resolutions of the filterbanks are reflected in the structures of the tonal residuals especially in speech pauses. In the second spectrogram the spectral peaks have all the same bandwidths and the same duration in time, while in the third and fourth spectrogram the bandwidths of the tonal residuals increase to higher frequency, but the durations in time decrease. Below 1 kHz the influence of the higher resolution becomes obvious. In this area the frequency structures are more detailed but the time resolution badly smears. In the high-frequency area the relations are reversed. Note that the amount of residual spectral peaks in higher frequency areas is significantly reduced using the wavelet filterbank.

In informal listening tests the residual noise produced by the nonuniform based enhancement systems was judged to be more pleasant than in the uniform solution. Furthermore the speech sounds more natural. In a direct comparison we give preference to the wavelet-based system.

In a second step we investigated more sophisticated spectral-subtraction rules with less residual noises. The application of the spectral-subtraction scheme of Ephraim and Malah leads to equivalent subjective results. Because the amount of residual noises is lower then in the case of Boll's procedure, the advantages of the non-uniform spectral analysis can be used to adjust the parameters to achieve less distortion of the speech signal.

6. CONCLUSION

The spectral subtraction method in conjunction with two filterbanks with nonuniform frequency bands is proposed. Informal listening tests stated the subjective preference of filterbanks with nonuniform bandwidths for spectral subtraction systems. The wavelet filterbank was judged to be superior to the allpass-transformed polyphase filterbank with the price of higher complexity in implementation.

7. REFERENCES

[1] Boll, Steven F. "Suppression of Acoustic Noise in Speech Using Spectral Subtraction", IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 27, pp 113-120, April 1979.

[2] Vary, P. "Noise Suppression by Spectral Magnitude Estimaton – Mechanism and Theoretical Limits –", EURASIP Signal Processing, Vol. 8, No. 4, pp 387-400, July 1985.

[3] Vary, P. "Ein Beitrag zur Kurzzeitspektralanalyse mit digitalen Systemen", Dissertation, Universität Erlangen-Nürnberg, 1978.

[4] Ephraim, Y. and Malah, D. "Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator", IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 32, pp 1109-1121, December 1984.

[5] Zwicker, E. "Psychoakustik", Springer Verlag, 1982.

[6] Shensa, M. J. "The Discrete Wavelet Transform: Wedding the Á Trous and Mallat Algorithms", IEEE Transactions on Signal Processing, Vol. 40, No. 10, pp 2464-2482, Oktober 1992.

[7] Daubechies, I.: *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1992.

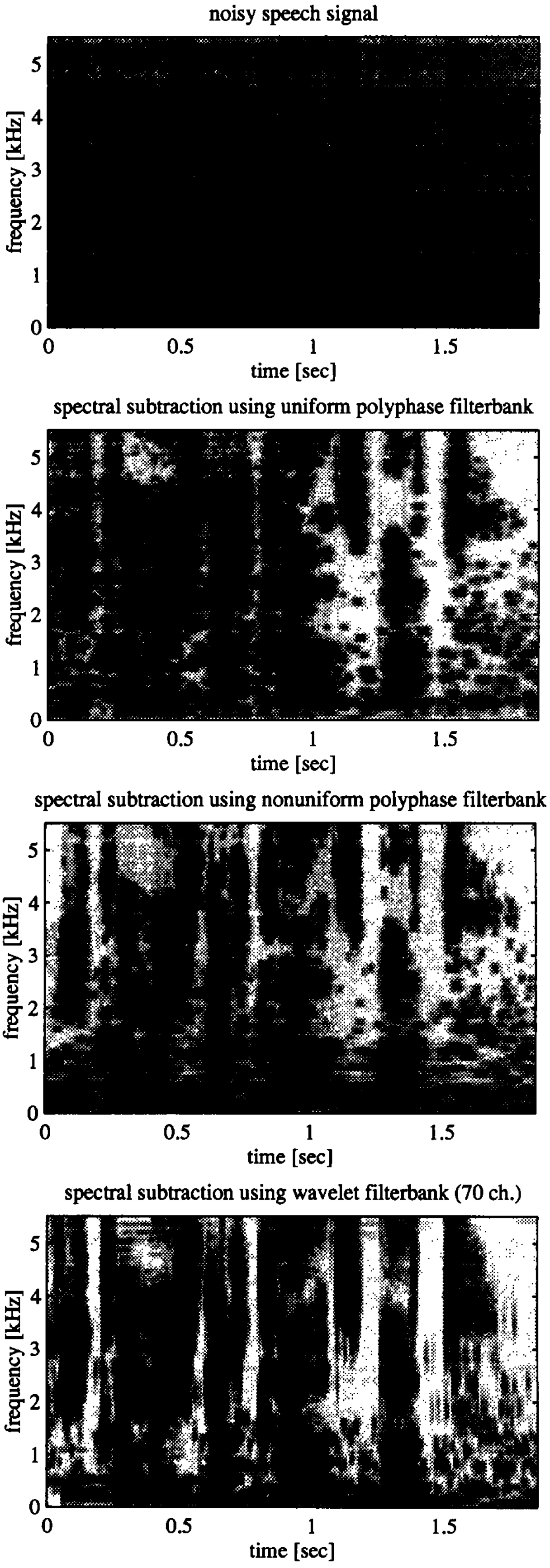

Figure 3: comparison of spectrograms.