Adaptive Comb Filtering in Speech Enhancement

Martin Vondra, Robert Vích

Institute of Photonics and Electronics
Academy of Sciences of the Czech Republic, Prague, Czech Republic
vondra@ufe.cz, vich@ufe.cz

Abstract

An improvement of the classical single-channel speech enhancement technique by comb filtering is presented in this paper. The comb filter is adaptive, tuned by the estimated pitch frequency and is applied in the spectral domain after speech enhancement based on spectral subtraction.

1. Introduction

There are many algorithms for enhancement of noise corrupted speech. These approaches can be roughly divided into single and multi-channel methods. A broad class of single channel methods is based on spectral subtraction algorithms [1]. In these methods the noise spectrum is estimated and then subtracted from the short-time spectrum of the noise corrupted speech. Various techniques have been developed for noise spectrum estimation either in speech pauses, or by spectral minima tracking [2], [3], etc. Also a large variety of spectral subtraction methods have been proposed [4], [5]. In spite of these sophisticated algorithms the enhanced speech contains a certain amount of residual noise and there remains considerable work to be done.

In this paper a novel two steps speech enhancement approach is presented. Additional suppression of the residual noise after the classical single-channel speech enhancement by spectral subtraction is performed by adaptive comb filtering applied in voiced frames only.

Specific properties of voiced speech signals, which can be considered as quasi harmonic signals, are exploited. The voiced speech signal $x(t)$ can be considered as a sum of sine waves, whose frequencies are integral multiples of the fundamental frequency F_0

$$x(t) = \sum_{k=1}^{N} C_k \sin(k\omega_0 + \varphi_k), \tag{1}$$

where ω_0 is the fundamental radian frequency, C_k and φ_k are the magnitude and phase of the k^{th} harmonic component. The number N is the assumed number of harmonics of the voiced speech signal.

A comb filter is a filter with multiple pass bands and stop bands. For transmitting only the harmonic components of the speech signal, the pass bands must be centered at multiples of the speech fundamental frequency, i.e. the frequency response of the comb filter has to be a periodic function with period equal to the fundamental frequency. Because voiced speech signals have time varying fundamental frequency, the comb filter for the enhancement of voiced speech has to be an adaptive filter tuned by the instantaneous fundamental frequency of the speech. It means that the comb filter vary from frame to frame.

A comb filter can be constructed by frequency transformation of a FIR or IIR prototype filter. Because almost all processing in speech enhancement algorithm based on spectral subtraction is performed in the spectral domain, it is appropriate to design and apply the comb filter in the spectral domain too.

2. Design of the comb filter

If we want to enhance the harmonic structure of voiced speech in the spectral domain, we design a comb filter in the spectral domain empirically. The requirements are narrow pass bands at multiples of the fundamental frequency and sufficient suppression in the stop bands. These requirements can be realized by placing of a rectangle pass band or cosine shaped function on each harmonic of the fundamental frequency F_0.

1.1. Rectangular pass bands

A typical magnitude frequency response of a comb filter is shown in Fig. 1. The spacing between the pass bands is given by the fundamental frequency F_0, the bandwidth $B<F_0/2$ is a chosen value dependent on the frame length and used window. The spectral maxima H_{max} and the spectral minima $H_{\min}$ are given by the required attenuation for the given signal to noise ratio (SNR). For the magnitude frequency response of a comb filter with rectangular path bands holds

$$\begin{aligned} H(f) &= H_{\max} \quad \text{for} \quad kF_0 - B/2 \le f \le kF_0 + B/2 \\ &= H_{\min} \quad \text{for} \quad (k-1)F_0 + B/2 < f < kF_0 - B/2, \end{aligned} \tag{2}$$

where $B<F_0/2$, $k=1,2,\ldots\text{fix}[F_s/2F_0]$ and F_s is the sampling frequency. The frequency $0 \le f \le F_s/2$ is sampled with the frequency step F_s/N_F, where N_F is the dimension of the applied FFT. The frequency response $H(f)$ for $F_s/2 < f \le F_s$ is given by $H(f) = H(F_s - f)$. The frequency response of such a comb filter can be efficiently generated by modulo operation.

The impulse response corresponding to the frequency response shown in Fig. 1 is depicted in Fig. 2.

1.2. Cosine shaped pass bands

For generation of a periodic magnitude frequency response of the comb filter we can also use the cosine function. For sharper spectral lobes we apply an exponential function with the cosine function in the exponent. The frequency response of such a comb filter is defined by

$$H(f) = \max\left\{ \frac{\exp[x\cos(2\pi f/F_0)]}{y},\ 10^{\mu/20} \right\}. \tag{3}$$

It holds $H_{\max} = e^x/y$ and $H_{\min} = 10^{\mu/20}$, $0 \le f \le F_s/2$.

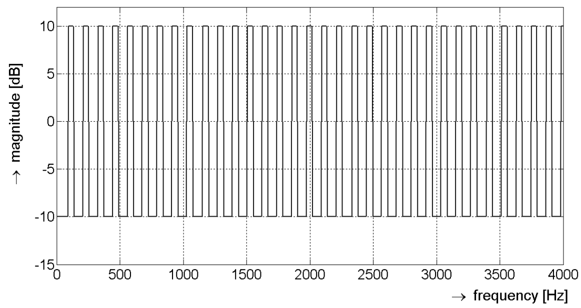

Figure 1: Frequency response of a comb filter with rectangular pass bands, where F_0 = 118Hz, B=47 Hz, H_{max}=3.2, H_{min}=1/ H_{max}.

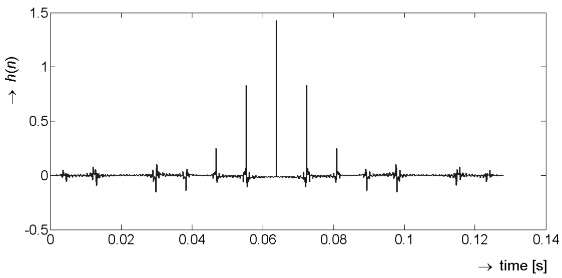

Figure 2: Impulse response of the comb filter shown in Fig. 1.

Parameters x and y are constants and they determine the bandwidth and magnitude maximum of each lobe. The minimum of (3) is limited to a certain number of decibels by the parameter μ =20logH_{min}. The 3dB bandwidth B of each lobe of this comb filter depends only on the parameter x and is expressed by

$$B = \frac{F_0}{\pi} \arccos\left(\frac{\ln(0.7079) + x}{x} \right). \tag{4}$$

The magnitude maximum of the of the lobes in the frequency response in decibels is given by

$$H_m = 8.686x - 20\log(y)\,. \tag{5}$$

The frequency response of such a comb filter for F_0 = 118Hz, x = 5, y = 50 and μ = –10dB is depicted in Fig. 3. The impulse response corresponding to the frequency response shown in Fig. 3 is depicted in Fig. 4.

3. Speech enhancement with comb filtering

Experiments have shown that comb filtering by itself is not sufficient to suppress noisy background in noise degraded speech signal. Further it is difficult to estimate the fundamental frequency for noisy speech. Therefore we have used comb filtering as post processing operation in speech enhancement, e.g. by spectral subtraction. The comb filter is constructed and applied in the frequency domain only for voiced frames after the classical spectral speech enhancement by spectral subtraction, see Fig. 5. For the construction of the comb filter we have to know the actual value of the speech fundamental frequency F_0. For its estimation it is appropriate to use a pitch determination algorithm also in the spectral domain [6]. If the spectrum after the classical spectral speech enhancement is identified as unvoiced, comb filtering is not applied.

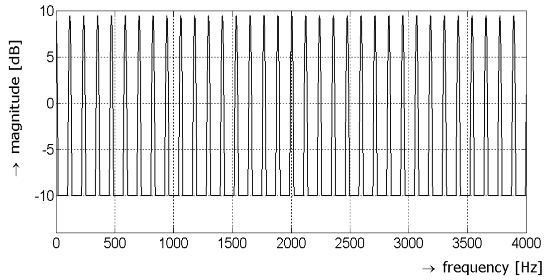

Figure 3: Frequency response of the comb filter designed according Eq. (2), where F_0 = 118Hz, x = 5, y = 50 and μ = *–10dB.*

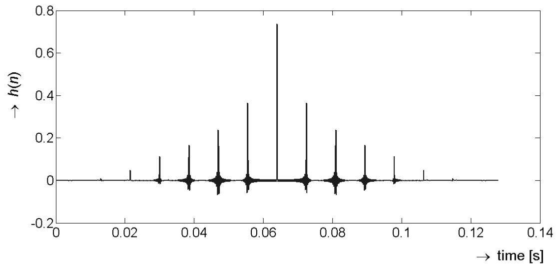

Figure 4. : Impulse response of the comb filter shown in Fig. 3.

4. Evaluation and Conclusion

The proposed method was tested and evaluated using noisy speech samples that were recorded in real environments. Only one microphone was used (no mixing of clean speech and noise). The signal from the microphone was sampled with F_s = 8 kHz and quantized with 16 bits. The SNR for various types of noise before enhancement and after enhancement by spectral subtraction and after enhancement by spectral subtraction and comb filtering are summarized in Table 1.

It can be seen from Table 1 that comb filtering offers a quite great SNR improvement. The SNR increase depends on the properties of the comb filter. Values in Table 1 were obtained for parameters H_{max}, H_{min}, x, y and μ cited in Fig. 1 and 3. Smaller bandwidth and higher maxima of the comb filter lobes yield greater SNR. But such a comb filter is more sensitive to precise estimate of F_0 and it has longer impulse response that causes distortion of the speech because of the time varying pitch period. A small value of the parameters H_{min} or μ can cause abrupt increase of additive noise in unvoiced part of speech, where comb filtering is not used.

Examples of spectrograms of noisy speech (car noise, 1st row in Table 1), enhanced speech by spectral subtraction and enhanced speech by spectral subtraction and comb filtering are shown in Fig. 6.

It can be stated that enhanced speech with comb filtering has smaller level of residual noise. A minor disadvantage is also a smaller level of unvoiced parts of speech, but by informal listening test it was observed that intelligibility is not decreased.

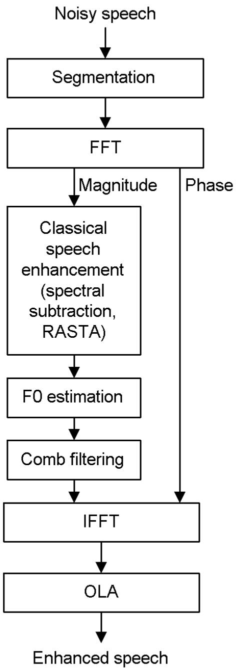

Figure 5: Block diagram of speech enhancement with comb filtering.

Table 1: SNR for noisy speech, for enhanced speech by spectral subtraction (SS) and for enhanced speech by spectral subtraction with comb filtering.

Type of noise	SNR [dB] Noisy speech	SNR [dB] SS	SNR [dB] SS + cos comb filter	SNR [dB] SS+rect. comb filter
Car	2.50	13.1	19.0	21.6
Vacuum cleaner	10.0	21.2	27.0	29.9
Shower	12.6	28.4	32.6	33.9
Electric drill	0.94	12.1	18.0	20.8

Informal listening test also showed that the subjective increase of speech quality for spectral subtraction with comb filtering in comparison with spectral subtraction itself does not correlate with the SNR values in Table 1.

5. Acknowledgements

This paper was prepared within the framework of the research project 102/07/1303 of the Grant Agency of CR and has been supported by the National research program "Information Society" of the Academy of Sciences of the Czech Republic, project number 1ET301710509.

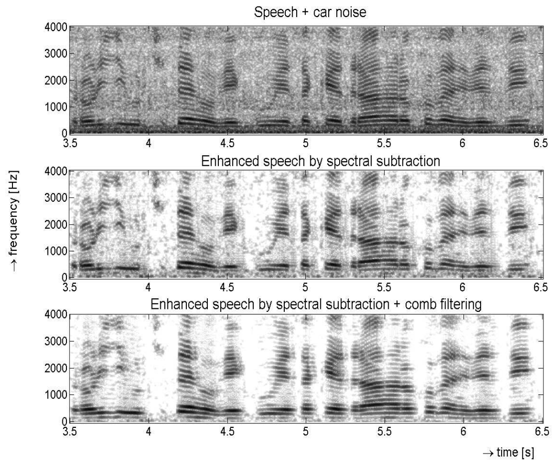

Figure 6: Spectrograms of noisy speech, enhanced speech by spectral subtraction and enhanced speech by spectral subtraction and comb filtering.

6. References

[1] Boll, S. F.: Suppressiomn of Acoustic Noise in Speech Using Spectral Subtraction, *IEEE Transactions on Acoustic, Speech, and Signal Processing*, vol. ASSP-27, No. 2, April 1979, pp. 113-120.

[2] Martin, R.: Noise power spectral density estimation based on optimal smoothing and minimum statistics, *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 5, July 2001, pp. 504-512,.

[3] Rangachari, S., Loizou, P. C.: A noise-estimation algorithm for highly non-stationary environments, *Speech Communication*, Vol. 48, pp. 220-231, 2006.

[4] Lim, J. S., Oppenheim, A. V.: Enhancement and Bandwidth Compression of Noisy Speech, *Proceedings of the IEEE*, vol. 67, No. 12, December 1979, pp. 1586-1604.

[5] Ephraim, Y., Malah, D.: Speech enhancement using a minimum mean-square error log-spectral amplitude estimator, *IEEE Trans. Acoust., Speech, Signal Processing*, April 1985, vol. ASSP-33, pp. 443–445.

[6] Schwarzenberg, M., Vích, R.: Robuste Grundfrequenz-bestimmung durch Korrelationsanalyse im Frequenz-bereich. In: Fortschritte der Akustik, DAGA 95, Saarbrücken, 1995, vol. II, pp. 1019-1022.