Adaptive Comb Filtering in Speech Enhancement by Spectral Subtraction

Martin Vondra, Robert Vích

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

Abstract: An improvement of the classical single-channel speech enhancement technique by comb filtering is presented in the paper. The comb filter is adaptive and tuned by the estimated pitch frequency. It is constructed empirically in the frequency domain using two approaches and is applied in voiced frames only also in frequency domain. The pitch frequency is estimated from enhanced speech in the frequency domain too. As classical speech enhancement technique, which is used prior comb filtering, the minimum mean-square error log-spectral amplitude estimator was applied. The obtained enhancement of speech corrupted by different noise sources is evaluated and compared for configuration with and without comb filtering by informal listening tests and using signal-to-noise ratio.

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 estimate the short-time speech magnitude spectrum based on a subtractive type of algorithm [1] or are based on multiplication of the short-time magnitude spectrum of noisy speech by a transfer function [2], [3]. In these methods the noise power spectrum has to be estimated. Various techniques have been developed for noise power spectrum estimation either in speech pauses by recursive averaging, in continuous speech by spectral minima tracking [4] or by minima controlled recursive averaging [5]. In spite of these sophisticated algorithms the enhanced speech contains a certain amount of residual and music noise and there remains considerable work to be done.

In some recently published papers [6], [7] authors try to improve speech enhancement by focusing on spectral harmonics of voiced speech. Similarly 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 is performed by adaptive comb filtering applied in voiced frames only. The application of adaptive comb filtering for speech enhancement is not a new approach. It was used in the time domain for example in [8] or [9], but in this paper we synthesize and apply the comb filter in the spectral domain.

Specific properties of voiced speech signals, which can be considered as quasi harmonic signals, are exploited. The voiced speech signal can be considered as a sum of sine waves, whose frequencies are integer multiples of the fundamental frequency F_0. 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 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 varies 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 can 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.

2.1 Rectangular pass bands

A typical magnitude frequency response of a comb filter with rectangular pass bands 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 pass bands holds

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

where $B < F_0/2$, $k = 1,2,\ldots\mathrm{fix}[F_s/2F_0]$ 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.

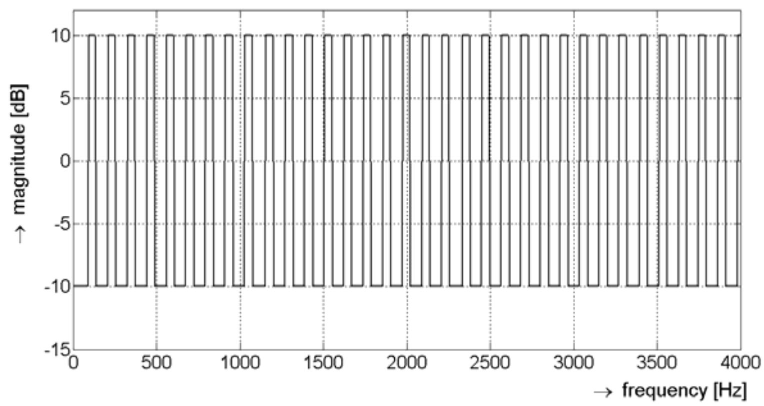

Figure 1 - Frequency response of a comb filter with rectangular pass bands, where F_0= 118Hz, B = 47Hz, $H_{\max}$= 3.2, $H_{\min}$= 1/ $H_{\max}$, F_s = 8 kHz.

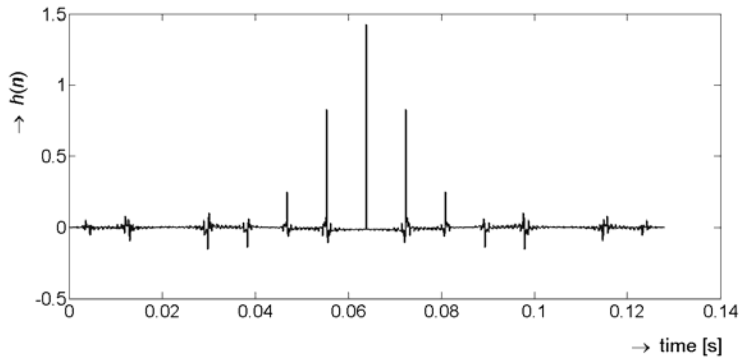

Figure 2 - Impulse response of the comb filter shown in Fig. 1.

2.2 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

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

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

Parameters x and y are constants and they determine the bandwidth and magnitude maximum of each lobe. The minimum of (2) is limited to a certain number of decibels by the parameter $\mu = 20\log H_{\min}$. The 3dB bandwidth B of each lobe of this comb filter depends only on the parameter x and is expressed by

$$B = \frac{F_0}{\pi}\arccos\left(\frac{\ln(0.7079) + x}{x}\right). \tag{3}$$

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

$$H_m = 8.686x - 20\log(y). \tag{4}$$

The frequency response of such a comb filter for $F_0 = 118$Hz, $x = 5$, $y = 50$, $\mu = -10$dB and $F_s = 8$kHz is depicted in Fig. 3. The impulse response corresponding to the frequency response shown in Fig. 3 is given in Fig. 4.

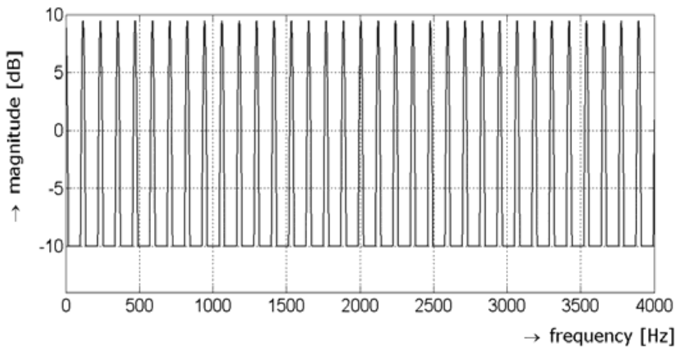

Figure 3 - Frequency response of the comb filter designed according Eq. (2), where F_0= 118Hz, $x = 5$, $y = 50$ and $\mu = -10$dB, $F_s = 8$ kHz.

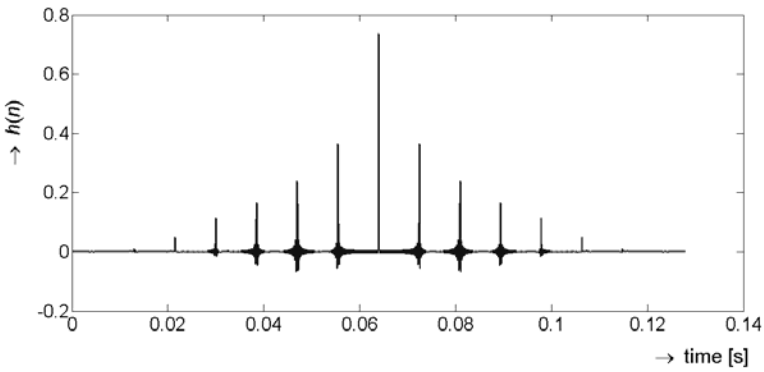

Figure 4 - Impulse response of the comb filter shown in Fig. 3.

3 Speech enhancement with comb filtering

We have used comb filtering as a post processing operation in speech enhancement e.g. by spectral subtraction. The comb filter is constructed only for voiced frames and applied in the frequency domain after the classical speech enhancement, see Fig. 5. Comb filtering is realized in the spectral domain by multiplication of the magnitude frequency response of the comb filter and of the short-time magnitude spectrum of enhanced speech. Phase remains unchanged by comb filtering. This operation brings some problems with the reconstruction of the time signal. First of all we must estimate the spectrum for a larger dimension N_F of the FFT to prevent cyclic convolution. For F_s = 8 kHz we use 24ms frames with 192 samples that are zero padded to N_F = 1024 samples for FFT computation. The frames must be overlapped. We use overlapping by half length of the frame, i.e. by 96 samples.

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 [10]. If the spectrum after the classical spectral speech enhancement is identified as unvoiced, comb filtering is not applied.

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 prior enhancement, after enhancement using Log Spectral Amplitude (LSA) speech estimator [3] and after the same enhancement 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 Figs. 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 a longer impulse response that causes distortion of the enhanced 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 parts of speech, where comb filtering is not used.

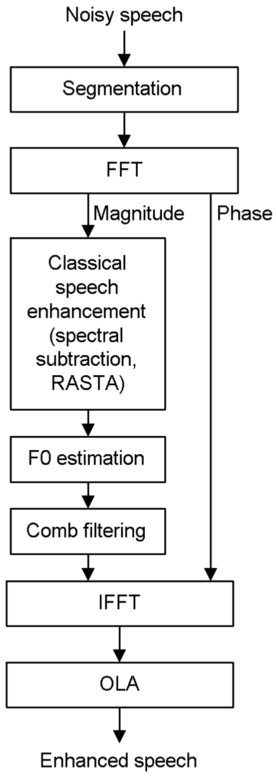

Figure 5 - Block diagram of speech enhancement with comb filtering

Type of noise	SNR [dB] Noisy speech	SNR [dB] LSA	SNR [dB] LSA + cos comb filter	SNR [dB] LSA+rect. comb filter
Car	2.50	14.3	20.2	22.7
Vacuum cleaner	10.0	24.0	29.6	32.3
Shower	12.6	25.8	28.8	29.9
Electric drill	0.94	12.0	17.5	20.0

Table 1 - SNR for noisy speech, for enhanced speech by LSA speech estimator and for enhanced speech by LSA speech estimator and comb filtering.

Examples of spectrograms of noisy speech (car noise, 1st row in Table 1), enhanced speech by LSA spectral estimator and enhanced speech by LSA spectral estimator and comb filtering are shown in Fig. 6. It can be stated that enhanced speech with comb filtering has smaller level of residual and music noise. A minor disadvantage is also a smaller level of unvoiced parts of speech, but by informal listening test it was observed that for described parameters of comb filters intelligibility is not decreased.

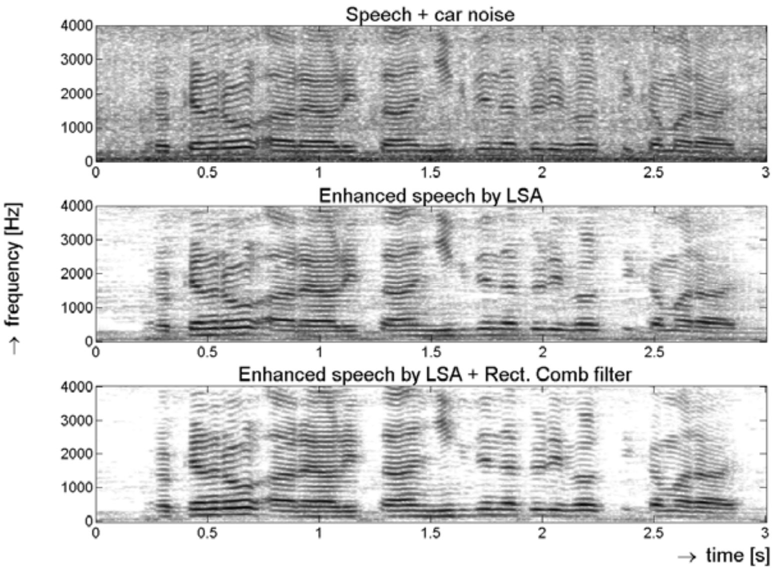

Figure 5 - Spectrograms of noisy speech, enhanced speech by LSA speech estimator and enhanced speech by LSA speech estimator and comb filtering (Czech sentence “Rock’n’roll a realita nikdy nebyli velcí kamarádi”).

Acknowledgements

This paper has been supported by the National research program "Information Society" of the Academy of Sciences of the Czech Republic, project number 1ET301710509.

References

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

[2] Ephraim, Y., Malah, D.: Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator. IEEE Transactions on Acoustic, Speech, and Signal Processing, ASSP-32, No. 6, December 1984, pp. 1109-1121.

[3] Ephraim, Y., Malah, D.: Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator. IEEE Transactions on Acoustic, Speech, and Signal Processing, April 1985, ASSP-33, pp. 443–445.

[4] Martin, R.: Spectral Subtraction Based on Minimum Statistics. In: Proc. 7th EUSIPCO’94, Edinburgh, U.K., Sept. 13-16, 1994, pp. 1182-1185.

[5] Cohen, I.: Noise Estimation by Minima Controlled Recursive Averaging for Robust Speech Enhancement. IEEE Signal Processing Letters, Vol. 9., No. 1, January 2002, pp. 12-15.

[6] Kim, H., Schwab, M., Moreau, N., Sikora, T.: Speech Enhancement of Noisy Speech Using Log-Spectral Amplitude Estimator and Harmonic Tunneling. In: Proc. IWAENC 2003, Kyoto, Japan, September 8-11, 2003.

[7] Zavarehei, E., Vaseghi, S., Yan, Q.: Noisy Speech Enhancement Using Harmonic Noise Model and Codebook-Based Post-Processing. IEEE Transactions on Audio, Speech, and Language Processing, ASLP-15, No. 4, May 2007, pp. 1194-1203.

[8] Lim, J., S., Oppenheim, A., V., Braida, L., D.: Evaluation of an Adaptive Comb Filtering Method for Enhancing Speech Degraded by White Noise Addition. IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-26, No. 4, August 1978, pp. 354-472.

[9] Hu, H., Kuo, F., Wang, H.: Supplementary Schemes to Spectral Subtraction for Speech Enhancement. Speech Communication, Vol 36, 2002, pp. 205-218.

[10] Schwarzenberg, M., Vích, R.: Robuste Grundfrequenzbestimmung durch Korrelationsanalyse im Frequenzbereich. In: Fortschritte der Akustik, DAGA 95, Saarbrücken, 1995, Vol. II, pp. 1019-1022.