EXTENDED SPECTRAL SUBTRACTION

Pavel Sovka & Petr Pollak & Jan Kybic
Czech Technical University, Faculty of Electrical Engineering
CTU FEL K331, Technicka 2, 166 27 Praha 6, Czech Republic
Tel: (+42 2) 2435 2291; Fax: (+42 2) 2431 0784
E-mail: [sovka,pollak]@feld.cvut.cz

ABSTRACT

This paper describes a new method for one channel noise suppression system which overcomes the typical disadvantage of one channel noise suppression algorithms - the impossibility of noise estimation during speech sequence. Our method is the combination of Wiener filtering and spectral subtraction. The noise can be successfully updated even during the speech sequences and that is why there is no need of the voice activity detector.

1 INTRODUCTION

The spectral subtraction offers the simple and computationally efficient tool for the suppression of an additive noise in a speech signal. This method has been extensively studied for almost twenty years. The research has been focused on higher degree of noise suppression, lower speech distortion, and less audible musical noise. The last requirement is important especially in the hand-free telephony application. But the main shortcoming of this method has not been overcome for a long time. It is the updating of the background noise characteristics estimation, especially during speech sequences.

2 SPECTRAL SUBTRACTION

The key idea of spectral subtraction is to estimate background noise and then to subtract this estimation from the noisy speech. The noise characteristics are usually updated during non-speech segments, i.e. the voice activity detector (VAD) is required to determine speech and non-speech sequences. Of course, the errors of this updating depends on the VAD quality, moreover, the updating cannot be provided during the speech activity segments. Two methods based on filter banks were published to overcome these shortcomings: Martin's [5] and Doblinger's [1]. These two methods are based on the observation that valleys of the short-time sub-band power estimate of noisy speech can be used by a long-time estimation of the background noise.

3 EXTENDED SPECTRAL SUBTRACTION

We suggest another approach based on the combination of spectral subtraction with iterative Wiener filtering. We refer this method as extended spectral subtraction [9]. The key feature of this method is the possibility to update the background noise estimation during speech activity segments.

We assume that we have a speech signal $s[n]$ corrupted by an additive noise $n[n]$. Noise is supposed to be uncorrelated with the speech and to be non-stationary. The rate of noise changes is relatively slower than the speech one. Then it is possible write

$$x[n] = s[n] + n[n]\,, \tag{1}$$

$$X(e^{j\theta}) = S(e^{j\theta}) + N(e^{j\theta}) \tag{2}$$

$$P_X(e^{j\theta}) = P_S(e^{j\theta}) + P_N(e^{j\theta})\,, \tag{3}$$

where $x[n]$ is discrete time representation of input signal, $X(e^{j\theta})$ its short-time discrete-time Fourier transform, $P_X(e^{j\theta})$ its power spectral density (PSD), etc. Whole algorithm is based on the estimation of $\widehat{N}(e^{j\theta})$ by the adaptive Wiener filter.

The frequently used approximation of the adaptive Wiener filter use the transfer function

$$H^2(e^{j\theta}) = \frac{|X(e^{j\theta})|^2 - |\overline{N}(e^{j\theta})|^2}{|X(e^{j\theta})|^2}\,, \tag{4}$$

where $|\overline{N}(e^{j\theta})|^2$ is the smoothed estimation of noise PSD performed in speech pauses by averaging of PSDs of input signal. Of course, VAD is needed in this case. The block scheme of this type of algorithms is on fig. 1.

Figure 1: Standard Wiener filter for speech enhancement.

There are some principal differences between our approach and the standard Wiener filter based algorithms.

Firstly, our approach belongs to the group of noise compensation structures, i.e. the Wiener filter is used to form the estimation of the input noise which is then subtracted from the input noisy speech. Secondly, the approximation of Wiener filter is done from the output signal not from the input one. That is the reason why the VAD is not needed. Thirdly, since the feed-back is used in the structure the algorithm can be viewed as some type of iterative Wiener filtering. The basic block scheme is on fig. 2.

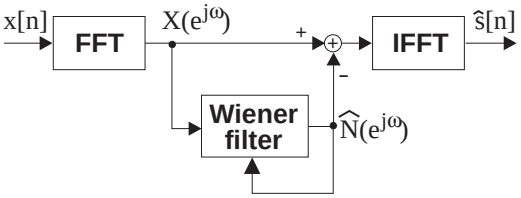

Figure 2: Principal structure of extended spectral subtraction.

The Wiener filter frequency response is set on the base of the preceeding spectral estimations according to following formula

$$H_n(e^{j\theta}) = \left(\frac{|\overline{N}_{n-1}(e^{j\theta})|^2}{|\overline{N}_{n-1}(e^{j\theta})|^2 + |\overline{S}_{n-1}(e^{j\theta})|^2} \right)^{1/2} \qquad (5)$$

where $|\overline{N}_{n-1}(e^{j\theta})|^2$ is the smoothed estimation of noise PSD and $|\overline{S}_{n-1}(e^{j\theta})|^2$ is the estimation of speech PSD.

The block scheme with more details is on fig. 3. The output of the Wiener filter is the estimation of noise which is currently subtracted from the input noisy speech and at the same time it is exponentially averaged. The smoothed noise estimation is then obtained according to

$$|\overline{N}_{n+1}(e^{j\theta})| = p \cdot |\overline{N}_n(e^{j\theta})| + (1-p) \cdot |\widehat{N}_n(e^{j\theta})| \,. \qquad (6)$$

The setting of the parameter p or the time constant

$$\tau \approx \frac{1}{1-p} \qquad (7)$$

strongly influences on the behaviour of the whole algorithm. If the rate of speech changes and the rate of noise changes are well separated then it is possible to set properly the parameter p and consequently the whole algorithm works well. Under this condition the slow changes appear in $|\overline{N}(e^{j\theta})|$ and in $|\widehat{N}(e^{j\theta})|$ however the faster ones (which are assumed to represent speech) appear in $|\overline{S}(e^{j\theta})|$.

The performance of this algorithm also depends on the setting other parameters. Typically, the short-time magnitude spectra can be used for Wiener filter approximation instead of PSDs, i.e.

$$H_n(e^{j\theta}) = \frac{|\overline{N}_{n-1}(e^{j\theta})|}{|\overline{N}_{n-1}(e^{j\theta})| + |\overline{S}_{n-1}(e^{j\theta})|} \,. \qquad (8)$$

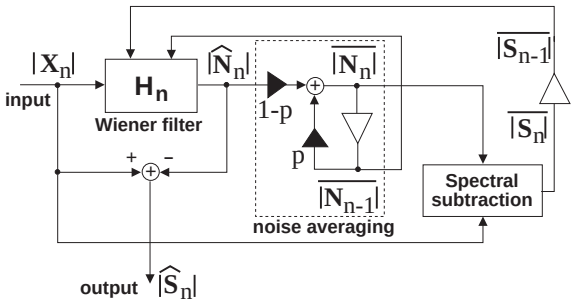

Figure 3: Block scheme of inner structure of extended spectral subtraction

This approximation yields the lower computation cost. Moreover, the speech distortion is less in this case, but unfortunately the noise suppression is less too. There is always some compromise in the choice of the proper Wiener filter approximation. The influence of the discussed parameter choice on short-time SNR is demonstrated on fig. 4.

Detail discussion is not possible in this short description. It is discussed in [9] and [6] together with the convergence analysis of the whole process.

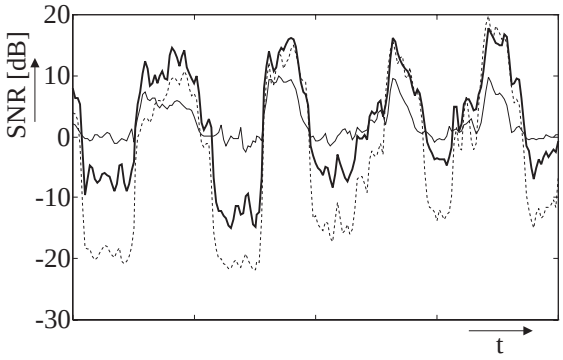

Figure 4: Local SNRs of input and output signals
Lines: dashed - noisy speech, solid - enhanced with a=2, solid bold - enhanced with a=1

4 VOICE ACTIVITY DETECTOR

Several VADs were published [2], [3] which are based on different principles. We have concentrated on cepstral VADs [7] which are very effective and yield low error rate. Commonly with speech enhancement we suggest now the modification of these VADs by including the possibility of noise cepstrum updating during speech activity. This approach enables to decrease the error rate of the VADs.

5 EXPERIMENTS AND RESULTS

Extended spectral subtraction was tested under real car noisy conditions. The different types of noises were cho-

Description	Algorithm specification
ffno	2-step spectral subtraction with FWR and no VAD
dobl	Doblinger's speech enhancement
mart	Martin's spectral subtraction
ext1	Extended spectral subtraction with $a = 2$ and $p = 0.8$
ext2	Extended spectral subtraction with $a = 2$ and $p = 0.93$
ext3	Extended spectral subtraction with $a = 1$ and $p = 0.93$

Table 1: Algorithm description

Algorithm	ffno	dobl	mart	ext1	ext2	ext3
nopaus.stac	1.35	6.01	7.41	6.07	6.86	5.29
nopaus.slow	0.33	6.32	7.27	6.10	6.87	5.30
nopaus.fast	-0.25	5.83	7.96	7.44	8.15	4.42

Table 2: Mean value of SSNRE for $\mathrm{SNR_{in}} = -5dB$

Algorithm	ffno	dobl	mart	ext1	ext2	ext3
nopaus.stac	1.88	37.10	55.10	36.98	47.11	27.98
nopaus.slow	3.34	40.49	53.03	37.33	47.26	28.16
nopaus.fast	3.20	37.17	64.63	56.87	68.16	20.86

Table 3: Variance of SSNRE for $\mathrm{SNR_{in}} = -5dB$

sen to demonstrate the ability to update the background noise spectrum during the speech activity. But in this case it is not possible to evaluate the classification criteria. We made some subjective listening tests by our research group.

To obtain the quantified results we realized the experiments with artificially mixed signals, i.e. the clean speech recorded in the car mixed with the noise recorded in the running car for specified SNR. This approach is possible because the noise in the running car can be considered as the additive one. For these experiments we computed the criteria described bellow.

5.1 Classification criteria

For the first tests we used the following criteria for speech enhancement classification based on the SNR: *segmental signal-to-noise ratio* - SSNR and SSNRE, i.e. the average of the short-time SNRs evaluated over speech frames only or its improvement respectively.

$$\mathrm{SSNRE} = \mathrm{SSNR_{out}} - \mathrm{SSNR_{in}}\,. \tag{9}$$

The short-time SNR is computed as

$$\mathrm{SNR}[i] = 10\log\frac{P_S[i]}{P_N[i]} = 10\log\frac{P_S[i]}{P_X[i] - P_S[i]}\,. \tag{10}$$

5.2 Results of experiments

Several algorithms, see tab. 1, were compared. We used two main types of speech signals - isolated words with pauses and speech without pauses - and three types of background noise - stationary noises, noises with relatively slow changes, and noises with fast changes.

When short words with many pauses were used then the performance of all methods was comparable. But without any VAD the performance of the spectral subtraction deteriorated while the performance of Doblinger's, Martin's, and extended spectral subtraction remained the same. The example of the response of the spectral subtraction and the extended spectral subtraction to the non-stationary noise can be seen on fig.5 - 7: the spectral subtraction produces the non-stationary residual noise with growing variance while the extended spectral subtraction produces the stationary residual noise with the same variance.

Typical results are summarized in tab. 2-3.

6 CONCLUSIONS

The new type of speech enhancement algorithm which is able to suppress the non-stationary noise in the speech without the need of VAD was described. The main idea is to use the difference between the rates of noise changes and the rate of speech changes. If these rates are different then no pauses in the speech are required. In comparison to Doblinger's and Martin's algorithms the noise estimation is got by Wiener filtering.

REFERENCES

[1] G. Doblinger. Computationally effecient speech enhancement by spectral minima tracking in subbands. In *EUROSPEECH'95 - Proceedings of the 4th European Conference on Speech Technology and Communication*, page 1513, Madrid, Spain, September 1995.

[2] J. A. Haigh and J. S. Mason. A voice activity detector based on cepstral analysis. In *EUROSPEECH'93 - Proceedings of the 3rd European*

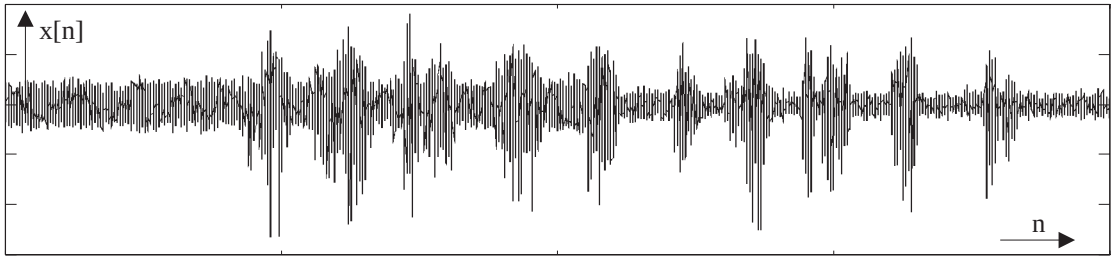

Figure 5: Input signal with non-stationary noise.

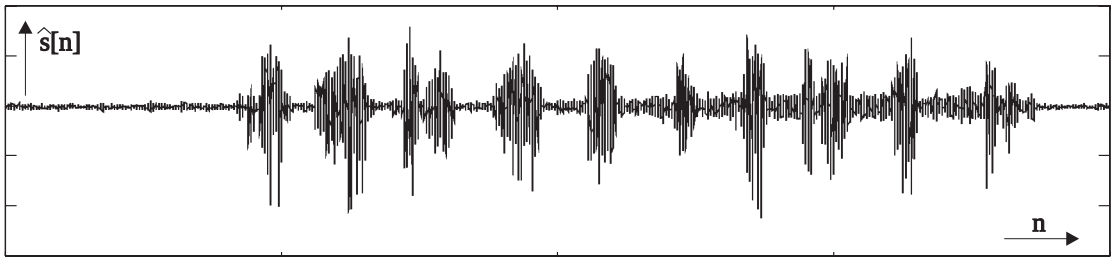

Figure 6: Output from standard spectral subtraction 'without' VAD.

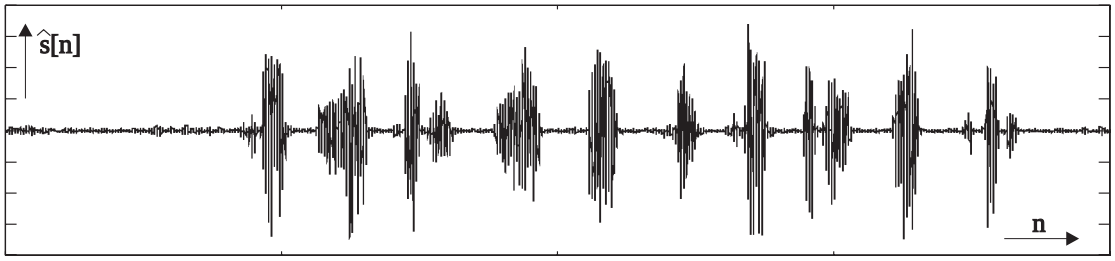

Figure 7: Output from extended spectral subtraction.

Conference on Speech, Communication, and Technology, pages 1103–1106, Berlin, September 1993.

[3] W. A. Harrison. Speech enhancement using multiply microphones. Technical Report #691, Massachusetts Institute of Technology, Lincoln laboratory, Group 24, November 1984.

[4] G. S. Kang and L. J. Fransen. Quality improvement of LPC-processed noisy speech by using spectral subtraction. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, ASSP-37(6):939–942, June 1989.

[5] R. Martin. Spectral subtraction based on minimum statistics. In *Proceedings of EUSIPCO-94 Seventh European Signal Processing Conference*, pages 1182–1185, Edinburgh, Scotland, U.K., September 1994.

[6] P. Pollák and J. Kybic. Porovnání jednovstupových metod potlačování aditivních šumů. Interní výzkumná zpráva R96-1, ČVUT - Elektrotechnická fakulta, Praha, 1996.

[7] P. Pollák, P. Sovka, and J. Uhlíř. Cepstral speech/pause detectors. In *Proceedings of IEEE Workshop on Nonlinear Signal and Image Processing*, Neos Marmaras, Greece, June 1995.

[8] P. Pollák, P. Sovka, and J. Uhlíř. Noise suppression system for a car. In *Proceedings of the 3rd European Conference on Speech, Communication, and Technology - EUROSPEECH'93*, pages 1073–1076, Berlin, September 1993.

[9] P. Sovka. Extended spectral subtraction - description and preliminary results. Research report R95-2, CTU - Faculty of Electrical Engineering, Prague, 1995.