ARCHIVES OF ACOUSTICS
35, 2, 203–212 (2010)

Adaptive Algorithms for Enhancement of Speech Subject to a High-Level Noise

Mariusz LATOS, Marek PAWEŁCZYK

Silesian University of Technology
Institute of Automatic Control
Akademicka 16, 44-100 Gliwice, Poland
e-mail: {Mariusz.Latos; Marek.Pawelczyk}@polsl.pl

(*received April 26, 2010; accepted May 14, 2010*)

There are many industrial environments which are exposed to a high-level noise, sometimes much higher than the level of speech. Verbal communication is then practically unfeasible. In order to increase the speech intelligibility, appropriate speech enhancement algorithms can be used. It is impossible to filter off the noise completely from the acquired signal by using a conventional filter, because of two reasons. First, the speech and the noise frequency contents are overlapping. Second, the noise properties are subject to change. The adaptive realisation of the Wiener-based approach can be, however, applied. Two structures are possible. One is the line enhancer, where the predictive realisation of the Wiener approach is used. The benefit of using this structure it that it does not require additional apparatus. The second structure takes advantage of the high level of noise. Under such condition, placing another microphone, even close to the primary one, can provide a reference signal well correlated with the noise disturbing the speech and lacking the information about the speech. Then, the classical Wiener filter can be used, to produce an estimate of the noise based on the reference signal. That noise estimate can be then subtracted from the disturbed speech. Both algorithms are verified, based on the data obtained from the real industrial environment. For laboratory experiments the G.R.A.S. artificial head and two microphones, one at back side of an earplug and another at the mouth are used.

Keywords: speech enhancement, adaptive system, line enhancer, LMS algorithm, high-level noise, nonstationary noise, earplug, active noise control.

1. Introduction

For increasing intelligibility of speech distorted by noise, a number of algorithms based on the general idea of spectral subtraction have been developed (Benesty *et al.*, 2008). For this algorithms a voice activity detector is usually used to distinguish between time frames, where the speech together with noise are present and those where the noise exists only. The frames with noise only allow to estimate its properties and then use them to eliminate the noise from the speech. If the signal-to-noise ratio is low, what is the case for industrial environments, the voice activity detection is poor and the remaining part of processing fails. Improperly subtracting the noise from a composite signal may even reduce speech intelligibility (Haykin, 1986; Widrow *et al.*, 1975; Widrow, Stearns, 1985).

This paper refers to real conditions existing in power plants, assembly lines, etc., where noise level may exceed 100 dB. Communication between people is of utmost interest for safety and job efficiency. Therefore, another approach to speech enhancement, which does not involve employment of voice activity detection and spectral subtraction will be used. It is based on filtering (Haykin, 1986; Michalczyk, 2004; Saxena *et al.*, 2008). Filters used for the above purpose can be fixed or adaptive. Knowledge about the signal and the noise is necessary to design fixed optimal filters. Adaptive filters have the ability to update their parameters on-line. Controlling the process in an adaptive system can in many cases help to accomplish the task with little risk of distorting the speech signal or increasing the output noise level. The filter-based approach is particularly effective if the noise to be reduced is periodic or narrowband, what is fortunately the case in most industrial environments due to working of rotating or reciprocating machines (Saxena *et al.*, 2008; Widrow *et al.*, 1975; Widrow, Stearns, 1985).

The problem of speech enhancement is a part of a larger project, which aims at designing a miniature personal active hearing protector supporting verbal communication among a group of users. The appropriate algorithm has to be chosen carefully to meet technical and operational requirements. It has to be relatively simple in implementation, fast and use small amount of the overall system resources. Such requirements are necessary to lower the power consumptions of the system and allow to use the protector for a working time without the necessity to replace or charge the battery.

2. Feedforward noise compensation

The classical realisation of the adaptive speech enhancer is presented in Fig. 1. A control filter W produces an estimate of the noise $d_1(n)$, disturbing the speech $s(n)$, based on the reference signal $d_2(n)$, correlated with the noise. An estimate of the speech signal $\hat{s}(n)$ is then obtained by subtracting the estimated noise from the disturbed speech signal. For success of this approach it is essential that the reference signal does not contain the speech signal. Otherwise, the speech would be reproduced partly by the filter and the overall enhancement effect would be reduced or the estimated speech could even be distorted. Such problem can be controlled to some extend if the noise were mostly composed of tonal components, by choosing a relatively small number of filter parameters (Widrow *et al.*, 1975; Widrow, Stearns, 1985). In case of environments where noise level is very high, placing of a cardioidal microphone at the back of the head or even at the ear completely suffices. The corrupted speech signal can be delayed by p samples before being processed to guarantee that the reference signal is acquired in advance and the control filter does not need to perform prediction.

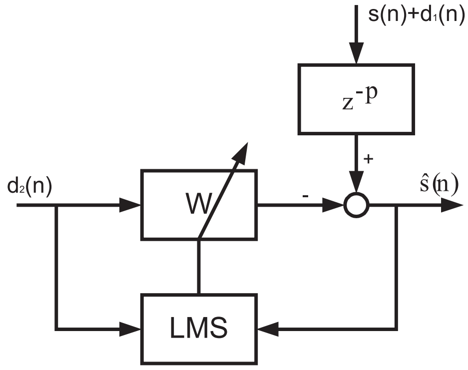


Fig. 1. Feedforward noise compensation.

A fixed parameter optimal Wiener filter would require precise modelling of signals or relevant paths (Elliott, 2001; Pawełczyk 2005). That is unfeasible because the noise is usually nonstationary. On the other hand, turning the head around with respect to the noise source results in dramatically different mutual dependences between those two signals. Therefore, an adaptive realisation is appreciated, where parameters $\underline{w}$ of the finite impulse response adaptive filter W of order N are updated with the Least Mean Square (LMS) algorithm (Elliott, 2001; Widrow, Stearns, 1985):

$$\underline{w}(n+1) = \underline{w}(n) + \frac{\mu}{\underline{d_2}^T(n)\underline{d_2}(n)} d_2(n)\hat{s}(n). \tag{1}$$

In (1) n is the current time instant, and μ is the convergence coefficient.

3. Prediction-based noise compensation

Application of the reference microphone discussed in the previous section increases the cost of the overall device and is sometimes ergonomically disadvantageous due to additional wires wrapped around the head. Therefore, it is justified to verify whether a system based on the speech recording microphone only can provide acceptable performance. For this purpose, the Wiener filter having a potential to predict signals is used in the system structure presented in Fig. 2. This structure is known in the literature as the line enhancer. The filter can successfully predict by k (chosen as relatively large) samples only deterministic or narrowband components of the recorded signal being the speech subject to noise. If the noise is of such a character, what is the case for most industrial conditions,

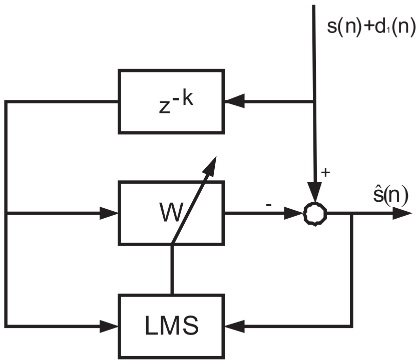


Fig. 2. Prediction-based noise compensation.

as explained earlier, the speech can be well reproduced since its spectral content is much richer (Widrow *et al.* 1975; Haykin, 1986).

4. Experiments

For experiments an active noise cancelling earplug with a microphone mounted at its back side has been tightly sealed to the G.R.A.S artificial ear (Fig. 3). That microphone has been used to provide a reference signal. The primary microphone acquiring the speech subject to noise is at the mouth.

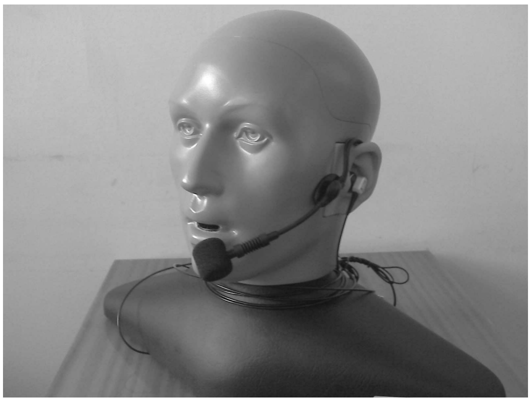


Fig. 3. G.R.A.S artificial head with mounted microphones.

The sampling frequency has been chosen as 8000 Hz, assuming that speech does not contain frequencies over 4 kHz. Real-world acoustic noise recorded in the Power Plant in Rybnik, Poland, has been used. Its PSD is presented in Fig. 6b by the dotted line. The LMS algorithm has been used for filter adaptation.

A set of experiments was conducted to obtain most appropriate values of control filter of order N, convergence coefficient μ for the LMS algorithm and the prediction horizon k for the line enhancer (Figs. 4, 5). Parameters resulting in the largest Signal-to-Noise Ratio (SNR) of the denoised speech were chosen. While tuning one parameter, the others were constant.

a) b)

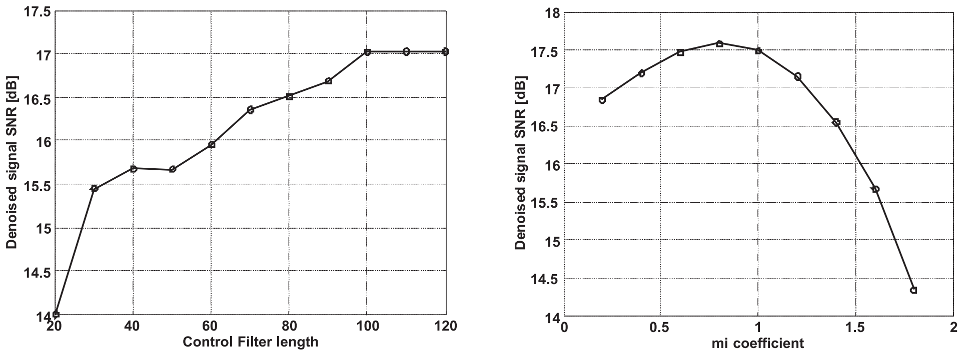


Fig. 4. Output estimate of the speech for the first system as a function of a) control filter order; b) convergence coefficient.

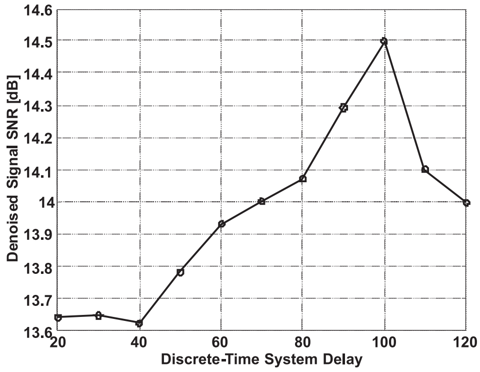


Fig. 5. Output estimate of the speech for the second system as a function of the prediction horizon, k.

In the next stage both systems were tested for the case of speech lacking at the primary input. For all experiments, the following parameters have been chosen as: $N = 100$, $\mu = 0.8$, $k = 100$ (for the line enhancer).

Assuming that the noise corrupting the speech is usually of low frequency, it is possible to support both systems by additional filtering of the estimated speech with a pass-band filter covering the frequencies 500–3200 Hz. After that operation, quality of the speech enhancer is increased. The obtained noise attenuation without output filtration is 7.5 dB and 6.9 dB, whereas with the filtration it is 15.2 dB and 13.4 dB, respectively, for the first and second system. Results are presented in Figs. 7–10. It is observed that the first system provides a better performance. The improvement is however not significant. Another option could

a) b)

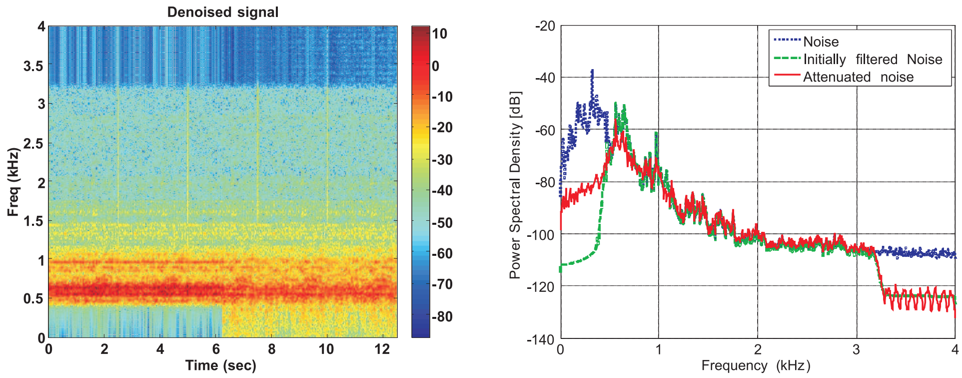


Fig. 6. a) Spectrogram demonstrating noise attenuation for the first system (system starts at the sample 50000), after initial band-pass filtering; b) Noise attenuation for the first system in the frequency domain.

a) b)

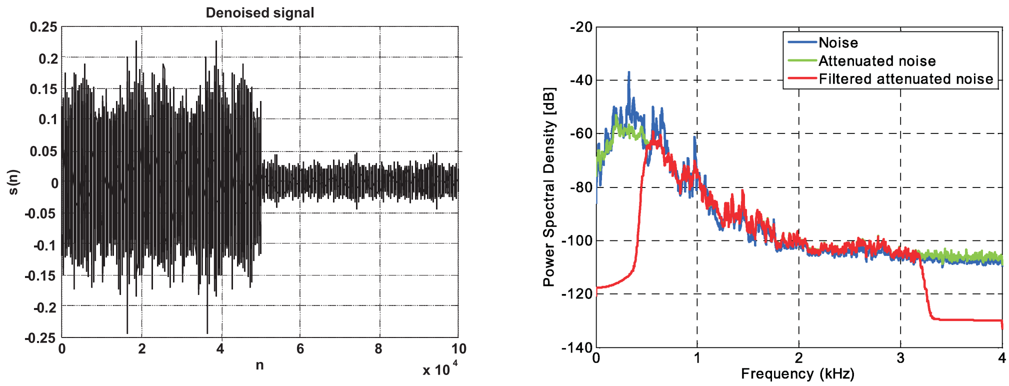


Fig. 7. Noise attenuation for the first system a) in the time domain (system starts at the sample 50000); b) in the frequency domain.

a) b)

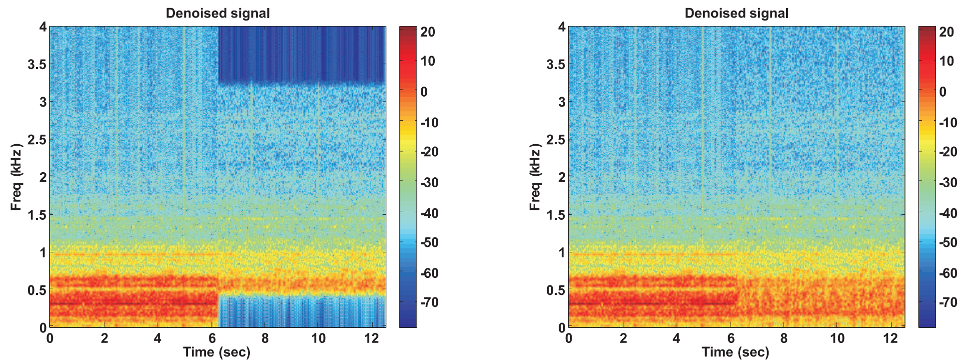


Fig. 8. Spectrograms demonstrating noise attenuation for the first system (system starts at the sample 50000) a) with the additional band-pass filtering; b) without the band-pass filtering.

a) b)

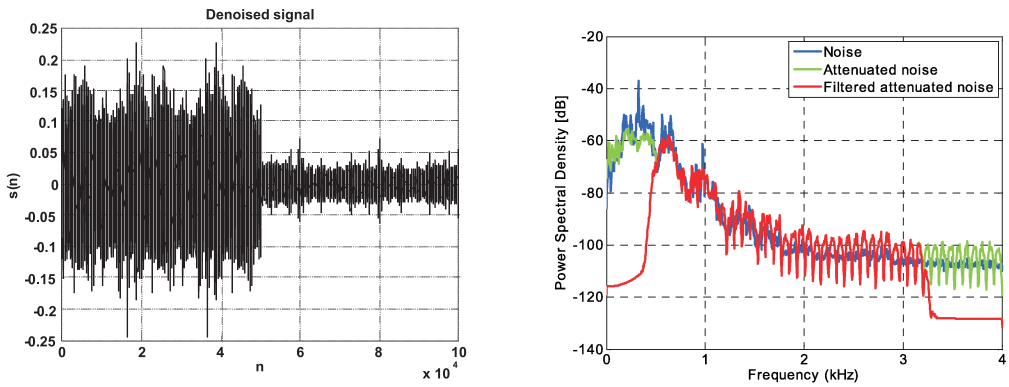


Fig. 9. Noise attenuation for the second system a) in the time domain (system starts at the sample 50000); b) in the frequency domain.

a) b)

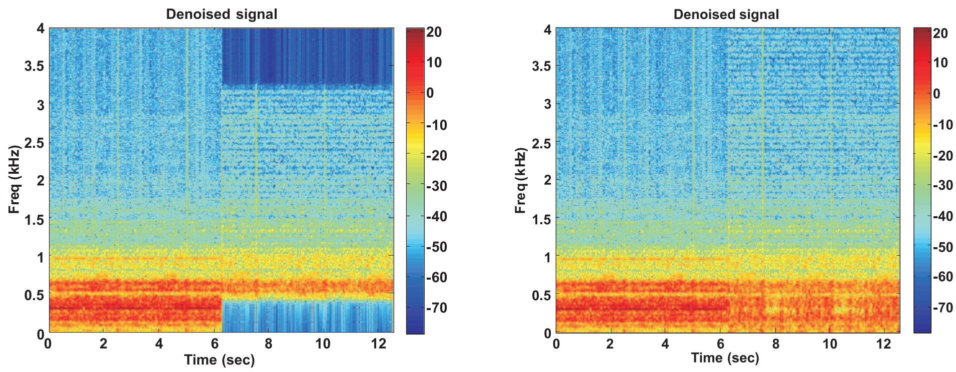
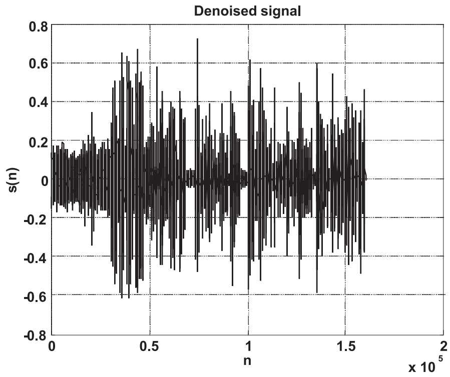


Fig. 10. Spectrograms demonstrating noise attenuation for the second system (system starts at the sample 50000) a) with the additional band-pass filtering; b) without the band-pass filtering.

be to apply the band-pass filter first and then use one of the adaptive structures. However, in that case the results are generally poorer, because of the averaging properties of the LMS algorithm. The noise attenuation obtained only through band-pass filtration is 9.3 dB, whereas initial filtration together with adaptive enhancer gives 15.0 dB for the first system (Fig. 6).

The last stage is to evaluate the quality of both systems in terms of speech enhancement. For the last set of experiments, primary input contained speech subject to noise characterised by SNR = 10.8 dB. The obtained enhanced speech signal SNR was equal to 16.0 dB and 14.1 dB without the band-pass filtering and 17.5 dB and 14.8 dB with that filtering, for both systems, respectively. The results are illustrated in Figs. 11–14. Similarly to the previous experiments, the approach employing the reference microphone yields better results for the selected type of noise.

a)



b)

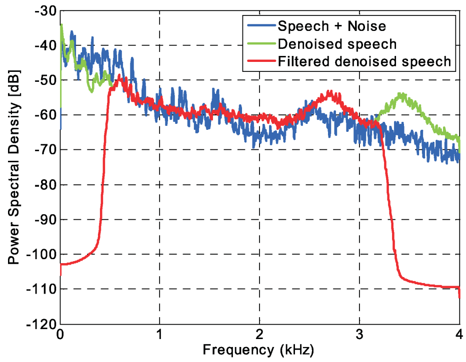
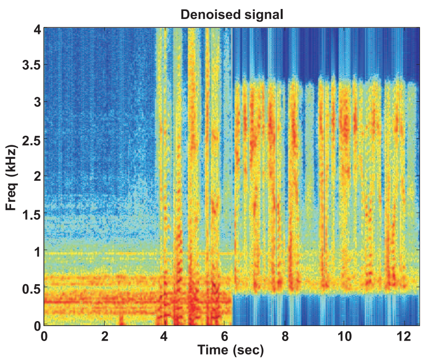


Fig. 11. Output estimate of the speech for the first system a) in the time domain (system starts at the sample 50000); b) in the frequency domain.

a)



b)

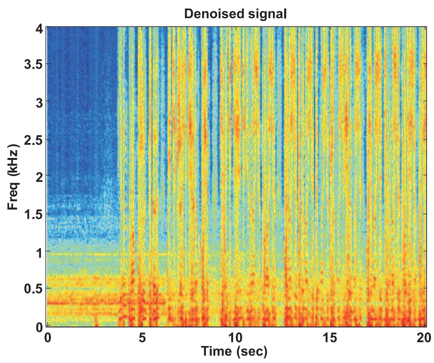
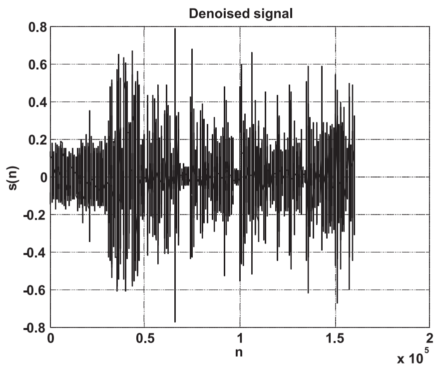


Fig. 12. Spectrograms demonstrating speech enhancement for the first system (system starts at the sample 50000) a) with the additional band-pass filtering; b) without the band-pass filtering.

a)



b)

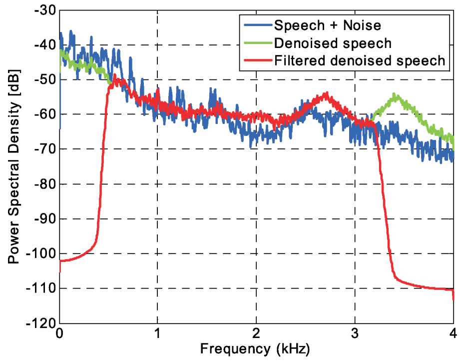
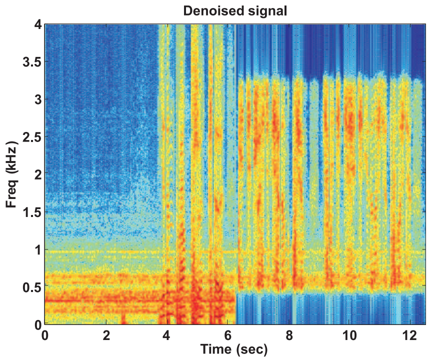


Fig. 13. Output estimate of the speech for the second system a) in the time domain (system starts at the sample 50000); b) in the frequency domain.

a) b)



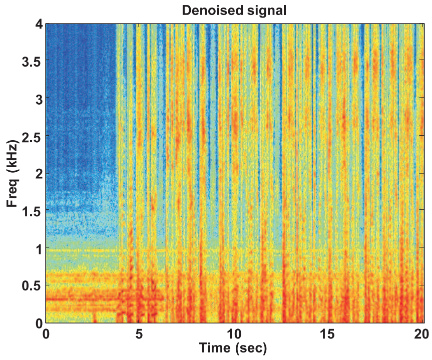


Fig. 14. Spectrograms demonstrating speech enhancement for the second system (system starts at the sample 50000) a) with the additional band-pass filtering; b) without the band-pass filtering.

5. Conclusions

In this paper adaptive Wiener-based approach to speech enhancement in the feedforward compensation and prediction structures has been considered. The LMS algorithm was used for adapting filter parameters. The principal advantages of the methods are its adaptive capability, relatively low output noise and low signal distortion. The adaptive capability allows the system to work properly, even when speech and noise characteristics are unknown and changing. Based on technological restrictions, these algorithm seem to be relevant for enhancing speech in environments exposed to a high-level noise. Supporting the systems with a fixed-parameter band-pass filtering significantly improves the performance. The system involving application of a microphone providing a signal correlated with the noise works significantly better for the type of disturbance under consideration.

Acknowledgment

The financial support of this research by the National Centre for Research and Development under the grant N R14 0031 06 is gratefully acknowledged.

References

1. Benesty J., Sondhi M.M., Huang Y., [Eds.] (2008), *Springer Handbook of Speech Processing*, Springer-Verlag, Berlin Heilderberg.
2. Elliott S. (2001), *Signal Processing for Active Control*, Academic Press, London.
3. Haykin S. (1986), *Adaptive Filter Theory*, Prentice Hall, New Jersey.
4. Michalczyk M.I. (2004), *Adaptive Control Algorithms for Three-Dimensional Zones of Quiet*, Jacek Skalmierski Computer Studio, Gliwice.

5. Pawełczyk M. (2005), *Feedback Control of Acoustic Noise at Desired Locations*, Silesian University of Technology, Gliwice.

6. Saxena G., Ganesan S., Das M. (2008), *Real-time Implementation of Adaptive Noise Cancelation*, IEEE International Conference on Electro/Information Technology, 431–436.

7. Widrow B., Glover J.R., McCool J.M., Kaunitz J., Williams C.S., Hearn R.H., Zeidler J.R., Dong E., Goodlin R.C. (1975), *Adaptive Noise Cancelling: Principles and Applications*, Proceedings of the IEEE, **63**, 12, 1692–1716.

8. Widrow B., Stearns S.D. (1985), *Adaptive Signal Processing*, Prentice Hall, New Jersey.