

# Enhanced estimation of power spectral density of noise using the wavelet transform

Petr Sysel<sup>1</sup> and Zdenek Smékal<sup>1</sup>

Department of Telecommunications, Brno University of Technology,  
Purkynova 118, 612 00 Brno, Czech Republic,  
sysel@feec.vutbr.cz

**Abstract.** In practice, different methods for enhancing speech hidden in noise are used but none of the available methods is universal; it is always designed for only a certain type of interference that is to be suppressed. Since enhancing speech masked in noise is of fundamental significance for further speech signal processing (subsequent recognition of speaker or type of language, compression, processing for transmission or storing, etc.), it is necessary to find a reliable method that would work even under considerable interference and will be modifiable for different types of interference and noise. The methods known to date can basically be divided into two large groups: single-channel methods and multi-channel methods. The basic problem of these methods lies in a rapid and precise method for estimating noise, on which the quality of enhancement method depends. If the noise is of stationary or quasi-stationary nature, its determination brings further difficulties. A method is proposed in the article for enhancing the estimation of power spectral density of noise using the wavelet transform.

**Keywords:** *Speech Enhancement, Power Spectral Density, Wavelet Transform Thresholding.*

## 1 Introduction

The wavelet analysis is a certain alternative to the Fourier representation for the analysis of short-term stationary real signals such as speech that is degraded by noise. If the noise is of non-stationary nature, then the greatest problem consists in estimating its power spectral density with sufficient frequency resolution. Two types of estimating the power spectral density are known: *non-parametric methods* and *parametric methods*. The best-known non-parametric methods include the Barlett method of periodogram averaging [1], the Blackman and Tukey method of periodogram smoothing [2], and the Welch method of averaging the modified periodograms [3]. Although the three methods have similar properties, the Welch method is the most widely used. These methods are called non-parametric because the parameters of the data being processed are not sought in advance. To yield a good estimate of the power spectral density the methods require the application of a long recording of data (at least  $10^4$  samples).

The periodogram is defined as follows:

$$P_{xx}(e^{j2\pi f}) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n] e^{-j2\pi f n} \right|^2 = \frac{1}{N} |X(e^{j2\pi f})|^2. \quad (1)$$

The function  $X(e^{j2\pi f})$  is the Fourier transform of discrete signal  $x[n]$ . It can be shown that the periodogram is an asymptotically unbiased estimate but its variance does not decrease towards zero for  $N \rightarrow \infty$ . This is to say that the periodogram itself is not a consistent estimate [4]. To be able to use the FFT, we must choose the discrete frequency values:

$$P_{xx}[e^{j2\pi f_k}] = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n] e^{-j2\pi f_k n} \right|^2 = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n] e^{-jk \frac{2\pi}{N} n} \right|^2, \quad (2)$$

$$k = 0, 1, \dots, N-1.$$

The periodogram is calculated at  $N$  frequency points  $f_k$ . For the comparison of the properties of non-parametric methods, the quality factor was proposed:

$$Q = \frac{\{E(P_{xx}[e^{j2\pi f_k}])\}^2}{\text{var}(P_{xx}[e^{j2\pi f_k}])}, \quad (3)$$

$$k = 0, 1, \dots, N-1.$$

A comparison of the non-parametric methods is given in Table 1. All the three methods yield consistent estimates of power spectral density. In the Bartlett method a rectangular window is used whose width of the main lobe in the frequency response when the maximum value drops by 3 dB is  $\Delta f = 0.9/M$ , where  $M$  is the length of partial sequences. In the Welch method a triangular window is used whose width of the main lobe in the frequency response when the maximum value drops by 3 dB is  $\Delta f = 0.28/M$ , where  $M$  is again the length of partial sequences and their overlap is 50%. In the Blackman-Tukey method, too, the triangular window is used and  $\Delta f = 0.64/M$ . As can be seen, in the Blackman-Tukey and the Welch methods the quality factor is higher than in the case of the Bartlett method. The differences are small though. What is important is that  $Q$  increases with increasing length of data  $N$ . This means that if for a defined value  $Q$  we want to increase the frequency resolution of estimate  $\Delta f$ , we must increase  $N$ , i.e. we need more data.

The main drawback of non-parametric methods is the fact that they assume zero values of the autocorrelation estimate  $r[m]$  for  $m \geq N$ . This assumption limits the frequency resolution of the periodogram. It is further assumed that the signal is periodic with period  $N$ , which is not true either. Since we have at our disposal only a sequence of finite length, there is aliasing in the spectrum.

There are, of course, also other methods that can extrapolate the values of autocorrelation for  $m \geq N$ . On the basis of the data analysed the parameters of the model are estimated (that is why they are called parametric methods), and using the model the properties of power spectral density are determined. Three

Method of PSD estimation	$Q$	Number of complex multiplications
Bartlett [1]	$1.11\Delta f N$	$\frac{N}{2} \log_2 \frac{0.9}{\Delta f}$
Blackman-Tukey [2]	$2.34\Delta f N$	$N \log_2 \frac{1.28}{\Delta f}$
Welch [3]	$1.39\Delta f N$	$N \log_2 \frac{1.28}{\Delta f}$

**Table 1.** Comparison of the quality of non-parametric methods according to [4].

types of model are known: *AR* (*Auto Regressive*), *MA* (*Moving Average*), and *ARMA*. From the three models, the AR model is the most frequently used. This is because it is well-suited to represent a spectrum with narrow peaks and then the Yule-Walker equation can be used to calculate the model coefficients. Wold [5] derived a theorem that says that any random process of the type of ARMA or MA can be represented uniquely using an AR model of infinite order.

The noise of blender was chosen for the comparison of different methods for estimating power spectral density. Blender noise is no typical random process since in addition to the random signal due to the friction between the blender content and blender parts it also contains periodic components due to rotor rotation. This noise was chosen because this type of noise is often encountered in speech enhancement. The majority of household appliances (vacuum cleaner, hair drier, etc.) and workshop machines (drilling, grinding, sawing and other machines) produce a similar type of noise. Fig. 1 gives a comparison of the estimates of power spectral density of blender noise obtained using a periodogram (full line) and a 15<sup>th</sup> order AR model (dashed line). As stated above, the periodogram is an asymptotically unbiased estimate but its variance may cause an error of as much as 100 % [6]. Moreover, it is inconsistent since it does not decrease with increasing signal length. This is shown in Fig. 1 by the large variance of values given by the full line. By contrast, the AR model provides a smoothed estimate of power spectral density. In the calculation, however, we encounter difficulties when estimating the order of the model. If a very low order of the model is used, we only obtain an estimate of the trend of power spectral density and lose the details, which in this case represent the maxima of the harmonics of periodic interference. If, on the contrary, the order of AR model is high, we obtain statistically unstable estimates with a large amount of false details in the spectrum. The variance of such an estimate will be similar to the estimate obtained using the periodogram. The choice of the order of the model is an important part of the estimation and depends on the statistical properties of the signal being processed. These properties are, however, in most cases also only estimated. For a short stationary signal the value of the order of the model can be chosen in the range:

$$0.05N \geq M \geq 0.2N, \quad (4)$$

where  $N$  is the signal length, and  $M$  is the order of the model [6]. For segments whose length  $N = 200$  samples the minimum order is  $M = 10$ . But for signals formed by a mixture of harmonic signals and noise this estimate mostly fails.

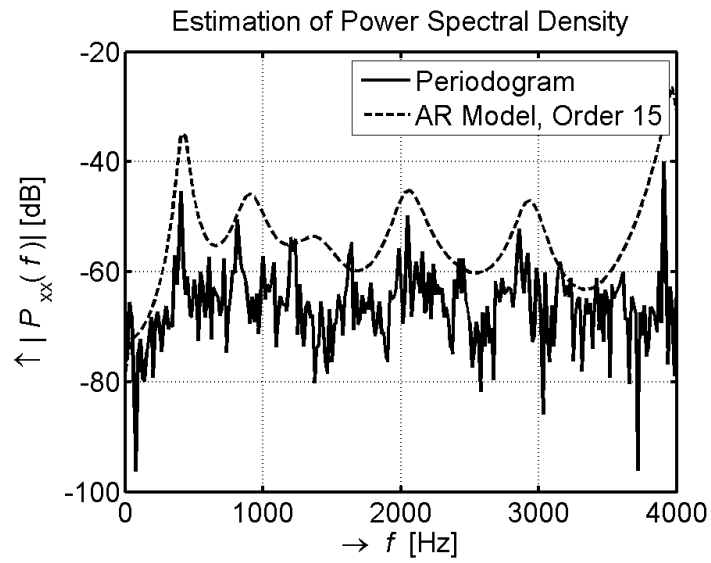
An estimate obtained using the periodogram can be made more precise via averaging the modified periodograms, which we obtain by dividing the signal into segments and weighting the latter by a weighting sequence. This approach is used in the Welch method for estimating the power spectral density [3]. In Fig. 2 we can see a comparison of the estimates of power spectral density obtained using the Welch method (with 13 segments weighted) and the AR model of 15<sup>th</sup> order. It is evident from the Figure that in comparison with the estimate obtained using the periodogram the Welch method yields a smoothed estimate, whose variance, moreover, decreases with increasing number of averaged segments (consistent estimate). Dividing the signal into segments naturally results in a reduced frequency resolution of the estimate, which shows up in less pronounced maxima that represent the components of periodic interference. In addition, in the case of non-stationary noise, periodograms may be averaged for segments that include interference of different statistical properties and thus also of different power spectral densities. This further reduces the accuracy of estimating the power spectral density.

## 2 Enhancing the estimate of power spectral density, using the wavelet transform

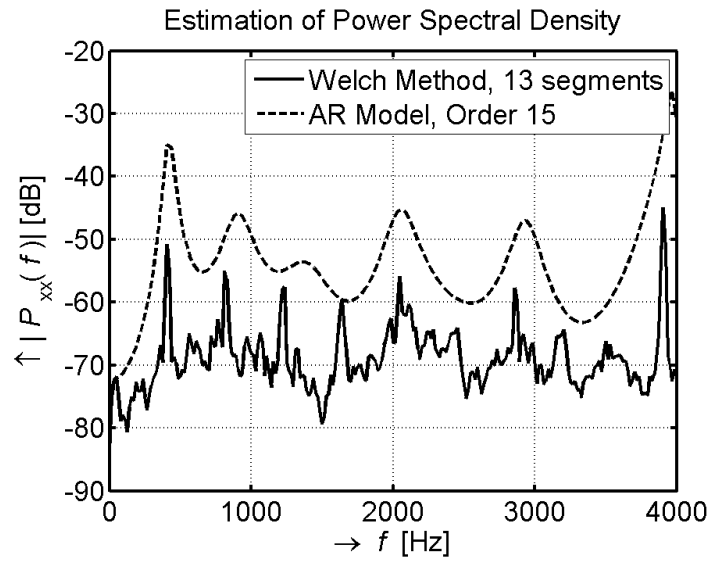
To enhance the estimate of power spectral density the wavelet transform can also be used. In [7] the non-parametric estimate of the logarithm of power spectral density is made more precise using the wavelet transform. With this method, statistically significant components of the estimate are obtained from thresholding different levels of wavelet decomposition and thus its non-negative values are ensured. The input signal is interfered with by additive non-Gaussian noise, and the wavelet coefficients of additive noise are assumed to be independent of the wavelet coefficients of power spectral density of useful signal. For the processing, type Daubechies wavelets of the 1<sup>st</sup>, 4<sup>th</sup>, 6<sup>th</sup> and 8<sup>th</sup> orders and type coiflet wavelets of the 2<sup>nd</sup> and 3<sup>rd</sup> orders were used. The method does not assume any preliminary knowledge of the type of noise. In [8] the authors endeavour to find a better estimate of power spectral density than the periodogram logarithm itself. To do this, they use the Welch method of modified periodograms, when they first average  $K$  periodograms that have first been multiplied by the spectral window. In [9] the method of wavelet transform thresholding is used to estimate noise, and this estimate is used to enhance speech by the spectral subtraction method.

Consider a stationary random process  $x[n]$ , which has a defined logarithm of power spectral density  $\ln G_{xx}(e^{j2\pi f})$ ,  $|f| \leq 0.5$ . As the function  $\ln G_{xx}(e^{j2\pi f})$  is periodic by frequency  $f$ , it can be expanded into a discrete Fourier series [4]:

$$\ln G_{xx}(e^{j2\pi f}) = \sum_{m=-\infty}^{\infty} v[m]e^{-j2\pi fm}. \quad (5)$$



**Fig. 1.** Estimation of power spectral density of blender noise, using the periodogram and the 15<sup>th</sup> order AR model



**Fig. 2.** Estimation of power spectral density of blender noise, using the Welch method of averaging modified periodograms for 13 segments, and the 15<sup>th</sup> order AR model.

Assuming that  $G_{xx}(e^{j2\pi f})$  is a real and even function by  $f$ , it holds that  $v[m] = v[-m]$ . The coefficients of the discrete Fourier series are:

$$v[m] = \int_{-0.5}^{0.5} \{\ln G_{xx}(e^{j2\pi f})\} e^{j2\pi f m} df, \quad (6)$$

$$m = 0, \pm 1, \pm 2, \dots$$

Discrete Fourier series coefficients  $v[m]$  are cepstral coefficients and the sequence  $v[m]$  is the cepstrum of autocorrelation sequence  $\gamma_{xx}[m]$ , where the Wiener-Khinchine relation holds:

$$G_{xx}[e^{j2\pi f}] = \sum_{m=-\infty}^{\infty} \gamma_{xx}[m] e^{-j2\pi f m}, \quad (7)$$

where it holds  $\gamma_{xx}[m] = E(x^*[n]x[n+m])$ , where the symbol  $*$  denotes a complex conjugate.

The estimate of autocorrelation sequence equals:

$$r_{xx}[m] = \frac{1}{2N+1} \sum_{n=-N}^N x^*[n]x[n+m], \quad (8)$$

and it holds:

$$\gamma_{xx}[m] = \lim_{N \rightarrow \infty} r_{xx}[m]. \quad (9)$$

The inverse equation to equation (5) has the form:

$$G_{xx}(e^{j2\pi f}) = e^{\left( \sum_{m=-\infty}^{\infty} v[m] e^{-j2\pi f m} \right)} = \sigma_w^2 H(e^{j2\pi f}) H(e^{-j2\pi f}), \quad (10)$$

where  $\sigma_w^2 = e^{v[0]}$  is the variance of white noise sequence  $w[n]$ . The transfer function  $H(z)$ ,  $z = e^{j2\pi f}$ , is the causal part of discrete Fourier series (10) and  $H(z^{-1})$  its non-causal part. In case the AR model is used, it holds for the transfer function:

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 + \sum_{i=1}^p a_i z^{-i}}. \quad (11)$$

The power spectral density will be obtained as follows:

$$G_{xx}(e^{j2\pi f}) = \sigma_w^2 |H(e^{j2\pi f})|^2, \quad (12)$$

where  $\sigma_w^2$  is the variance of white noise sequence  $w[n]$ , for which it holds:

$$\sigma_w^2 = E(|w[n]|^2). \quad (13)$$

The periodogram  $P_{xx}(e^{j2\pi f})$ , as the estimate of power spectral density of noise  $G_{xx}(e^{j2\pi f})$ , is defined as follows:

$$P_{xx}(e^{j2\pi f}) = \sum_{m=-(N-1)}^{N-1} r_{xx}[m]e^{-j2\pi fm} = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n]e^{-j2\pi fn} \right|^2. \quad (14)$$

Using the periodogram we can determine the power spectral density as:

$$G_{xx}(e^{j2\pi f}) = \ln G_{xx}(e^{j2\pi f}) + \epsilon(e^{j2\pi f}) + \gamma. \quad (15)$$

If we assume noise to be a Gaussian random process, then the logarithm of periodogram can be modelled as:

$$\ln P_{xx}(e^{j2\pi f}) = \ln G_{xx}(e^{j2\pi f}) + \epsilon(e^{j2\pi f}) + \gamma, \quad (16)$$

where  $\epsilon(e^{j2\pi f})$  is a random process with probability distribution  $\chi_2^2$  with two degrees of freedom, and  $\gamma \approx 0.57721$  is the Euler-Mascheroni constant [7, 11]. The random process  $\epsilon(e^{j2\pi f})$ , which is responsible for the periodogram variance, can be removed via thresholding the wavelet transform coefficients. The coefficients of the discrete wavelet transform with discrete time  $C_{j,m}[k]$  will be calculated according to the relation:

$$C_{j,m}[k] = \sum_{k=0}^{N-1} (\ln P_{xx}[k] - \gamma) \psi_{j,m}[k], \quad (17)$$

where  $P_{xx}[k]$  are samples of the periodogram of the implementation of a random process of length  $2N = 2^{M+1}$  obtained by the discrete Fourier transform, and the base functions  $\psi_{j,m}[k]$  are derived by a time shift  $j = 0, 1, \dots, 2^{m-1}$  and dilatation with the scale  $m = 0, 1, \dots, \log_2 N$  of a single mother function  $\psi[k]$  according to the relation:

$$\psi_{j,m}[k] = \frac{1}{\sqrt{2^m}} \psi \left[ \frac{k}{2^m} - j \right]. \quad (18)$$

The transform is linear and therefore the coefficients will represent the sum of the coefficients representing the sought power spectral density  $g_{j,m}[k]$  and the coefficients representing the noise  $e_{j,m}[k]$ :

$$C_{j,m}[k] = g_{j,m}[k] + e_{j,m}[k]. \quad (19)$$

As the random processes  $\epsilon[k]$  are independent and the transform is orthogonal, the coefficients  $e_{j,m}[k]$  will be non-correlated. At the same time, however, they are not independent since their probability distribution is independent of the shift  $j$  but is dependent on the scale  $m$ . But according to the central limiting theorem their probability distribution converges with increasing  $m \rightarrow \infty$  to the Gaussian normal distribution. The coefficients are modified via soft thresholding according to the relation:

$$C_{j,m}^{(s)}[k] = \text{sgn}(C_{j,m}[k]) \max(0, |C_{j,m}[k]| - \lambda), \quad (20)$$

where  $\lambda$  is the threshold value chosen. After that the smoothed estimate of power spectral density of noise  $\hat{G}_{xx}[k]$  is obtained (via the inverse discrete wavelet transform) in the form:

$$\ln \hat{G}_{xx}[k] = \frac{1}{N} \sum_{m=0}^M \sum_{j=0}^{2^m-1} C_{j,m}^{(s)}[k] \psi_{j,m}[k], \quad (21)$$

$$k = 0, 1, \dots, N-1.$$

With regard to the asymptotically normal distribution of coefficients  $e_{j,m}[k]$  the threshold value  $\lambda$  can be determined using the universal thresholding proposed by Johnstone, et al., [10] in the form:

$$\lambda = \sigma \sqrt{2 \log N}, \quad (22)$$

where  $\sigma$  is the standard deviation of noise, and  $N$  is the length of data. In [11] an optimum determination of the threshold in dependence on the scale is proposed. If the scale is large, then the threshold equals:

$$\lambda_m = \alpha_m \ln N, \quad (23)$$

where the constants  $\alpha_m$  are given in Table 2 and  $N$  is the length of data. If the scale is small,  $m \ll M-1$ , the threshold will be determined according to the relation:

$$\lambda_m = \frac{\pi}{\sqrt{3}} \sqrt{\ln N}. \quad (24)$$

The blender noise, which was used in testing the estimations using the AR model, the Welch method and the periodogram, was also used in the estimation via thresholding the coefficients of the wavelet transform of the periodogram. A comparison of this estimation with that using the AR model is shown in Fig. 3. It is evident that the estimate using the AR model is more smoothed but has a smaller frequency resolution than the estimate obtained via thresholding the

scale $m$	$\alpha_m$	scale $m$	$\alpha_m$
$M-1$	1.29	$M-6$	0.54
$M-2$	1.09	$M-7$	0.46
$M-3$	0.92	$M-8$	0.39
$M-4$	0.77	$M-9$	0.32
$M-5$	0.66	$M-10$	0.27

**Table 2.** Values of constant  $\alpha_m$  for determining the threshold when thresholding the wavelet transform coefficients.



wavelet transform coefficients. On the contrary, the estimate obtained via thresholding the wavelet transform coefficients is less smoothed but the pronounced peaks in the power spectrum, which represent the harmonic components of the blender motor interference, are localized in the frequency more easily than in the case of AR models. This is also evident from a comparison of the method of thresholding the wavelet transform coefficients with the Welch method of thresholding modified periodograms in Fig. 4.

### 3 Experimental results

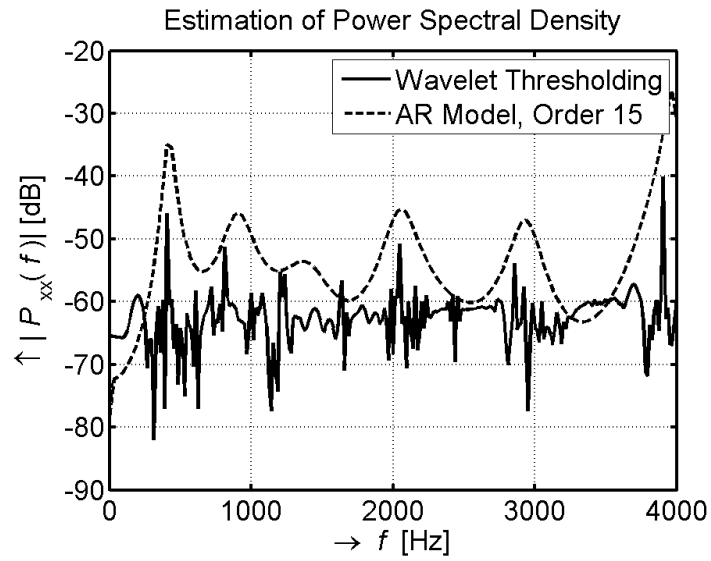
The proposed method of power spectral density estimation was tested on speech enhancement by spectral subtraction [12] of actual recordings of speech signal interfered with by different types of noise. The speech signal was interfered with by the noise of blender and vacuum cleaner, which have the character of wide-band noise almost approximating white noise. In the testing, the speech signal was also exposed to interference by noise from a drilling machine and a Ford Transit, which on the contrary have the character of narrow-band noise. In view of the fact that the recordings were made in a real environment and it was impossible to obtain pure speech signal without interference, the estimation of the quality was performed using the signal-to-noise ratio determined segmentwise according to the relation:

$$\text{SNR}_{\text{seg}} = 10 \log_{10} \frac{R_s}{R_\nu}. \quad (25)$$

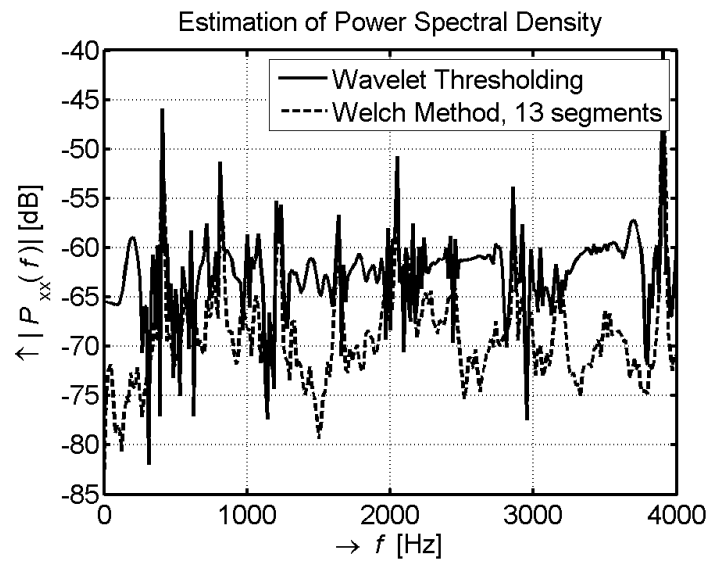
The power of signal  $R_s$  is determined from a segment containing the speech signal while the power of interference  $R_\nu$  is estimated from a segment that does not contain the speech signal.

A noisy speech signal was enhanced using the proposed modified method of spectral subtraction. For comparison, the speech was also enhanced by the method of spectral subtraction with the power spectral density of interference being estimated using the Welch method. For enhancement, the RASTA method was also applied. The values of SNR of the original signal and of the signal reconstructed by the individual methods are given in Table 3.

Compared to the original spectral subtraction method and the RASTA method there was a marked improvement in the SNR in the case of interference of wideband character - vacuum cleaner or blender. Less good results are obtained in the case of noise of narrow-band character, where additionally much depends on the position of narrow-band noise. If noise is in the same frequency band as speech, e.g. drilling machine, the modified method of spectral subtraction exhibits an improvement of SNR which is only a little lower than the improvement in the original method of spectral subtraction or the RASTA method. If the position of narrow-band noise is outside the speech frequency band (low-frequency noise of the Ford Transit engine), then the modified method is comparable with the other methods.



**Fig. 3.** Estimation of power spectral density of blender noise, via thresholding the wavelet transform coefficients of periodogram and via the AR model of 15<sup>th</sup> order.



**Fig. 4.** Estimation of power spectral density of blender noise, via thresholding the wavelet transform coefficients of periodogram and via the Welch method of averaging modified periodograms for 13 segments.

Estimation of signal-to-noise ratio $SNR_{seg}$ [dB]				
Method	Original	Modified	Spectral	RASTA
Type of noise	signal	method	subtraction	method
Vacuum cleaner	12	22	16	16
Blender	5	17	12	13
Drill	0	13	17	14
Ford Transit	3	20	18	17

**Table 3.** Estimation of signal-to-noise ratio for four different types of noise and three different methods of speech signal enhancement.

## 4 Conclusions

A new method for enhancing the estimation of power spectral density of noise via thresholding the wavelet transform coefficients was proposed. It led to a reduced variance in the estimate of power spectral density of noise. The smoothed estimate was then used to enhance the speech signal hidden in noise, using the spectral subtraction method. Thanks to the lower variance in the estimate of power spectral density of noise there was also a lower variance in the estimate of the power spectrum of speech signal and a reduction in the signal reconstruction error. Judging by the experiments made, the method is in the first place suitable for noise that is of wideband nature - e.g. shower, vacuum cleaner, and the like. Listening tests showed that thanks to the lower noise variance the musical noise was suppressed which was the by-product of the existing method of spectral subtraction. The estimation of power spectral density of noise via thresholding the wavelet transform coefficients can also be used in other single-channel methods of speech enhancement.

## Acknowledgements

This work was supported within the framework of project No 102/07/1303 of the Grant Agency of the Czech Republic and the National Research Project "Information Society" No 1ET301710509.

## References

1. Bartlett, M.S.: Smoothing Periodograms from Time Series with Continuous Spectra. *Nature* (London), Vol. 161 (May 1948), 686-687.

2. Blackman, R. B., Tukey, J. W.: The Measurement of Power Spectra. Dover, New York (1958).
3. Welch, P. D.: The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method of Time Averaging over Short Modified Periodograms. In IEEE Trans. Audio and Electroacoustics, Vol. AU-15 (June 1967), 70-73.
4. Proakis, J.G., Manolakis, D.G.: Digital Signal Processing—Principles, Algorithms and Applications. Third Edition. Prentice Hall, New Jersey (2006).
5. Wold, H.: Study in the Analysis of Stationary Time Series. Reprinted by Almqvist & Wiksell, Stockholm (1954).
6. Uhlír, J., Sovka, P.: Digital Signal Processing. Prague: CVUT Publishing(1995). (In Czech)
7. Moulin, P.: Wavelet Thresholding Techniques for Power Spectrum Estimation. In IEEE Transactions on Signal Processing, Vol. 42, No 11 (1994), 3126-3136.
8. Walden, A.T., Percival, D.B., McCoy E.J.: Spectrum Estimation by Wavelet Thresholding of Multitaper Estimators. In IEEE Transactions on Signal Processing. Vol. 46, No. 2 (1998), 3153-3165.
9. Sysel, P.: Wiener Filtering with Spectrum Estimation by Wavelet Transform. In Proceedings of International Conference on “Trends in Communications”. Bratislava, Slovakia:(2001), 471-474.
10. Donoho, D. L., Johnstone, L. M., Kerkyacharian, G., Picard, D.: Wavelet shrinkage: asymptotic (with discussion). J. R. Statist. Soc. B, Vol. 57, (1995), 301-369.
11. Vidakovic, B.: Statistical Modeling by Wavelets. John Wiley & Sons, Inc. Publication, New York (1999).
12. Benesty, J., Makino, S., Chen, J.: Speech Enhancement. Springer, Berlin (2005).