Estimation of Power Spectral Density using Wavelet Thresholding

PETR SYSEL
Brno University of Technology
Department of Telecommunications
Purkynova 464/118, 612 00 Brno
CZECH REPUBLIC

JIRI MISUREC
Brno University of Technology
Department of Telecommunications
Purkynova 464/118, 612 00 Brno
CZECH REPUBLIC

Abstract: The basic problem of the single-channel speech enhancement methods lies in a rapid and precise method for estimating noise, on which the quality of enhancement method depends. The paper describes a new method of power spectral density estimation using wavelet transform in spectral domain. To reduce periodogram variance the proposed method use the procedure of thresholding the wavelet coefficients of a periodogram. Then the smoothed estimate of power spectral density of noise is obtained using the inverse discrete wavelet transform.

Key–Words: Speech Enhancement, Power Spectral Density, Periodogram, Wavelet Transform Thresholding

1 Introduction

Most of the speech enhancement methods use estimation of noise and interference characteristics. Thus the notion of power spectral density is introduced, which defines the density of total noise energy of a random signal in dependence on frequency.

By the Wiener-Khinchin relation [3, 8] it can be derived that the power spectral density \( G_{xx} [k] \) of a stationary random process can be obtained by the discrete Fourier transform of the autocorrelation sequence

\[
G_{xx} [k] = \sum_{m=-\infty}^{\infty} \gamma_{xx} [m] e^{-j2\pi km}, \quad (1)
\]

where \( \gamma_{xx} [m] \) is the autocorrelation sequence of stationary random process, for which it holds

\[
\gamma_{xx} [m] = \sum_{n=-\infty}^{\infty} x[n + m]x[n]. \quad (2)
\]

The stationary random process denotes a random process whose mean value and variance are independent of time \( n \) while the autocorrelation sequence \( \gamma_{xx} [m] \) depends only on lags \( m \).

The autocorrelation sequence in formula (1) is theoretically infinite and in practice it must be replaced by an estimate from one or a few process realizations of finite length. On the other hand, this causes distortion and the power spectral density obtained is only an estimate \( \hat{G}_{xx} [k] \) of theoretical power spectral density \( G_{xx} [k] \).

The estimation bias (systematic error) \( B \{ \hat{G}_{xx} [k] \} \) is defined as the difference between the theoretical power spectral density \( G_{xx} [k] \) and the mean value of power spectral density estimation \( \hat{G}_{xx} [k] \) [8]

\[
B \{ \hat{G}_{xx} [k] \} = G_{xx} [k] - E \{ \hat{G}_{xx} [k] \}. \quad (3)
\]

Estimation is unbiased when the mean value of estimation is equal to the theoretical power spectral density

\[
E \{ \hat{G}_{xx} [k] \} = G_{xx} [k]. \quad (4)
\]

Estimation variance \( D \{ \hat{G}_{xx} [k] \} \) is defined as the mean square value of the difference between the estimate and the mean value of estimate [8]

\[
D \{ \hat{G}_{xx} [k] \} = E \left\{ \left( \hat{G}_{xx} [k] - E \{ \hat{G}_{xx} [k] \} \right)^2 \right\}. \quad (5)
\]

Thus the estimation error can be defined by normalized mean square error [8]

\[
\epsilon_c^2 = \epsilon_b^2 + \epsilon_r^2, \quad (6)
\]

where \( \epsilon_b^2 \) is the normalized mean square error due to bias, and \( \epsilon_r^2 \) is the normalized mean square error due to variance. Both errors are defined as

\[
\epsilon_b^2 = \left( \frac{B \{ \hat{G}_{xx} [k] \}}{G_{xx} [k]} \right)^2, \quad (7)
\]

\[
\epsilon_r^2 = \frac{D \{ \hat{G}_{xx} [k] \}}{G_{xx}^2 [k]} \quad (8)
\]


2 Power spectral density estimation methods

2.1 Non-parametric methods

Non-parametric methods of power spectral density estimation use the discrete Fourier transform. The autocorrelation sequence can be estimated by the relation

\[ r_{xx}[m] = \frac{1}{N} \sum_{n=0}^{N-m-1} x[n+m]x[n], \quad 0 \leq m < N, \]

\[ r_{xx}[m] = \frac{1}{N} \sum_{n=-|m|}^{N-1} x[n+m]x[n], \quad -N < m < 0, \]

where \( N \) is the signal length. The estimation used is biased but it asymptotically approximates the theoretical autocorrelation sequence

\[ \lim_{N \to \infty} r_{xx}[m] = \gamma_{xx}[m]. \] (10)

Moreover, estimation variance converges to zero as \( N \to \infty \), therefore \( r_{xx}[m] \) is a consistent estimate [3].

If we substitute (9) into (1), an estimate of power spectral density of an ergodic random process, so called periodogram, can be obtained

\[ P_{xx}[k] = \sum_{m=-(N-1)}^{N-1} r_{xx}[m] e^{-j2\pi km} = \frac{1}{N} |X[k]|^2. \] (11)

However, the periodogram is also a biased estimate but asymptotically it is an unbiased estimate

\[ \lim_{N \to \infty} \mathbb{E}\{P_{xx}[k]\} = G_{xx}[k]. \] (12)

When the random process is a Gaussian random process, the estimation variance is shown to be [3]

\[ \text{D}\{P_{xx}[k]\} = G_{xx}^2[k] \left[ 1 + \left( \frac{\sin 2\pi \frac{k}{N}}{N \sin 2\pi \frac{k}{2N}} \right)^2 \right]. \] (13)

Hence the estimation variance is approaching the square of power spectral density

\[ \lim_{N \to \infty} \text{D}\{P_{xx}[k]\} = G_{xx}^2[k] \] (14)
as \( N \to \infty \). Therefore estimation using the periodogram is an inconsistent estimation. Moreover, estimation variance achieves the value of the square of power spectral density and the normalized mean square error due to variance

\[ c^2_t = \lim_{N \to \infty} \frac{\text{D}\{G_{xx}[k]\}}{G_{xx}^2[k]} = \frac{G_{xx}^2[k]}{G_{xx}^2[k]} = 1 \] (15)

achieves the value 1, i.e. 100%. For this reason estimation using the periodogram is improper.

One of the methods of smoothing the power spectral density estimation is the mean value of a few periodograms, which are obtained from a few sufficiently long random process realizations. [3, 8] If we have at our disposal only one random process realization, then we subdivide an \( N \)-point signal into \( K \) segments. The effect of subdividing is reduced frequency resolution by a factor \( K \). For each segment, the periodogram is computed

\[ P_{xx}^{(i)}[k] = \frac{1}{M} \left| \sum_{n=0}^{M-1} x_i[n]e^{-j2\pi kn} \right|^2, \]

where \( M = \left\lfloor \frac{N}{K} \right\rfloor \) is the length of the each segment. Then the power spectral density estimation is equal to the mean value of partial periodograms

\[ P_{xx}^{B}[k] = \frac{1}{K} \sum_{i=1}^{K-1} P_{xx}^{(i)}[k]. \] (16)

This method is called the Bartlett method [3]. The power spectral density estimation by the Bartlett method is again an asymptotically unbiased estimation. However, the estimation variance is reduced by a factor \( K \) [3]

\[ \text{D}\{P_{xx}^{B}[k]\} = \frac{1}{K} G_{xx}^2[k] \left[ 1 + \left( \frac{\sin 2\pi \frac{k}{N}}{N \sin 2\pi \frac{k}{2N}} \right)^2 \right]. \] (17)

After substituting into (8) the normalized mean square error due to variance is shown to be reduced by a factor \( K \) too.

Another modification of averaging periodograms is proposed by Welch. The fundamentals are the same as in the Bartlett method but Welch proposed segment overlapping and weighting segments by a chosen window. The Welch method can reduce the estimation variance while preserving sufficient frequency resolution. [3]

2.2 Parametric methods

Unlike non-parametric methods the parametric methods emulate a model for the generation of signals that is constructed from a number of parameters estimated from a short realization of the random process observed. The modelling approach makes it possible to extrapolate the value of the random process outside the realization and also to extrapolate the value of the autocorrelation sequence for lags \(|m| \geq N\). An advantage is that these methods
provide a better frequency resolution than non-parametric methods and their frequency resolution is almost independent of the length of random process realization. [3]

The discrete stationary random process \( x[n] \) with unknown power spectral density \( G_{xx}[k] \) can be transformed into white noise \( w[n] \) when sequence \( x[n] \) is filtered by a linear digital filter with a transfer function \( 1/H(z) \) that is inverse to the transfer function (20). This filter is then called whitening filter. If it is possible to find out the transfer function of whitening filter then the power spectral density is shown to be

\[
G_{xx}(e^{j2\pi f}) = \sigma_w^2 H(e^{j2\pi f}) \overline{H(e^{j2\pi f})},
\]

where

\[
\sigma_w^2 = e^{v[0]}
\]

is the variance of white noise,

\[
H(z) = \exp \left( \sum_{m=1}^{\infty} v[m] z^{-m} \right) = \sum_{i=0}^{r} b_i z^{-i}, |z| > r_1, r_1 < 1,
\]

is the transfer function of the model for the observed random process generation, and \( v[m] \) are the cepstral coefficients of the autocorrelation sequence \( \gamma_{xx}[m] \).

2.3 Enhancing the estimate of power spectral density using the wavelet transform

Other methods of power spectral density estimation can be based on thresholding the wavelet coefficient of periodogram [9, 7, 4, 5]. Consider a stationary random process \( x[n] \), which has a defined logarithm of power spectral density \( \ln G_{xx}(e^{j2\pi f}) \), \(|f| \leq 0.5\). If we assume noise to be a Gaussian random process, then the logarithm of periodogram can be modelled as

\[
\ln P_{xx}[k] = \ln G_{xx}(e^{j2\pi f}) + \zeta(e^{j2\pi f}) + \gamma,
\]

where \( \zeta(e^{j2\pi f}) \) 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. [2, 9]

The random process, 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

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

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 \( j = 0, 1, \ldots, 2^m - 1 \) and dilatation with the scale \( m = 0, 1, \ldots, \log_2 N \) of a single mother function \( \psi[k] \) according to the relation

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

The transformation 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 \( e_{j,m}[k] \)

\[
C_{j,m}[k] = g_{j,m}[k] + e_{j,m}[k].
\]

As the random processes \( \zeta[k] \) are independent and the transformation 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 to the Gaussian normal distribution with \( m \to \infty \) [2]. The coefficients are modified via soft thresholding according to the relation

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

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

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

\[
k = 0, 1, \ldots, N - 1.
\]

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 Donoho [1] in the form

\[
\lambda = \sigma \sqrt{2 \log N},
\]

where \( \sigma \) is the standard deviation of noise, and \( N \) is the half of data length. In [9] an optimum determination of the threshold in dependence on the scale is proposed. If the scale is large, then the threshold equals

\[
\lambda_m = \alpha_m \ln N,
\]
where the constants \( \alpha_m \) are as given in Table 1 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

\[
\lambda_m = \frac{\pi}{\sqrt{3}} \sqrt{\ln N}.
\]

<table>
<thead>
<tr>
<th>scale ( m )</th>
<th>( \alpha_m )</th>
<th>scale ( m )</th>
<th>( \alpha_m )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M - 1 )</td>
<td>1.29</td>
<td>( M - 6 )</td>
<td>0.54</td>
</tr>
<tr>
<td>( M - 2 )</td>
<td>1.09</td>
<td>( M - 7 )</td>
<td>0.46</td>
</tr>
<tr>
<td>( M - 3 )</td>
<td>0.92</td>
<td>( M - 8 )</td>
<td>0.39</td>
</tr>
<tr>
<td>( M - 4 )</td>
<td>0.77</td>
<td>( M - 9 )</td>
<td>0.32</td>
</tr>
<tr>
<td>( M - 5 )</td>
<td>0.65</td>
<td>( M - 10 )</td>
<td>0.27</td>
</tr>
</tbody>
</table>

Table 1: Values of constant \( \alpha_m \) for determining the threshold when thresholding the wavelet transform coefficients.

![Figure 1: Power spectral density estimation using periodogram, AR model 10\(^{th}\) order, and wavelet thresholding of circular saw noise.](image)

3 Experiments

The proposed method of power spectral density estimation using wavelet thresholding was tested on speech enhancement by spectral subtraction [6] of actual recordings of speech signal interfered with by different types of noise. The block diagram of the modified spectral subtraction is in the Fig. 2. [4, 5]

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 and the AR model with order \( s = 20 \). The speech signal was interfered with by the noise of blender and vacuum cleaner, which have the character of wideband 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. Approximately 60 recordings from an actual environment with various signal-to-noise ratio and with various types of interference (drill, vacuum cleaner, vehicles, blender, shower, etc.) are processed. [6] 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 segment-wise according to the relation

\[
\Delta \kappa = 10 \log_{10} \frac{R_s}{R_v}.
\]

The power of signal \( R_s \) is determined from a segment containing the speech signal while the power of interference \( R_v \) is estimated from a segment that does not contain the speech signal.

The resulting signal-to-noise ratio and improvement in the signal-to-noise ratio were monitored in all enhanced recordings. The mean value and the variance of improvement of signal-to-noise ratio were calculated. The results are summarized in Table 2.

4 Conclusion

A new method for enhancing the estimation of power spectral density of noise using thresholding the wavelet transform coefficients was demonstrated. It led to a reduced variance in the estimate of power spectral density of noise. The accuracy of power spectral density estimation can greatly affect the results of the methods that use it. This can be demonstrated on speech enhancement by the spectral subtraction method with various types of interference and various signal-to-noise ratios. 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.

Although the mean value of improvement in the signal-to-noise ratio changes very little depending on the method of power spectral density estimation used, marked differences can be seen in the variance of improvement in the signal-to-noise ratio. It can be said that methods of power spectral density estimation with a smaller variance of estimation (thresholding of wavelet coefficient of periodogram) provide a smaller variance of the improvement of signal-to-noise ratio. Hence these methods are more suitable because the
Figure 2: Modified spectral subtraction method using the wavelet transform to estimate the power spectral density of noise.

Table 2: Results of spectral subtraction using various power spectral estimation methods.

<table>
<thead>
<tr>
<th>Power spectral density estimation by</th>
<th>Improvement of signal-to-noise ratio $\Delta \kappa$ (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Minimum</td>
</tr>
<tr>
<td>Welch method</td>
<td>2,72</td>
</tr>
<tr>
<td>wavelet thresholding</td>
<td>3,52</td>
</tr>
<tr>
<td>AR model $s = 20$</td>
<td>3,69</td>
</tr>
</tbody>
</table>

improvement in the signal-to-noise ratio is less dependent on the type of interference or on the input signal-to-noise ratio. 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.

Acknowledgements: This work was supported within the framework of the project No 102/07/1303 of the Grant Agency of the Czech Republic and the National Research Project “Information Societ” No 1ET301710509 by the Academy of Sciences of the Czech Republic. This work is also supported in part by the Ministry of Education, Youth and Sports of the Czech Republic under research program No. MSM0021630513.

References: