SPEECH ENHANCEMENT BASED ON THE STRUCTURE OF NOISE POWER SPECTRAL DENSITY

Chengshi Zheng, Yi Zhou, Xiaohu Hu, and Xiaodong Li

Key Laboratory of Noise and Vibration Research, Institute of Acoustics, Chinese Academy of Sciences
No.21, Bei-Si-huan-Xi Road, 100190, Beijing, China
email: {cszheng,yizhou,hxh,lxd}@mail.ioa.ac.cn

ABSTRACT

This paper proposes an adaptive averaging periodogram (AAP) spectral estimator based on the structure of noise power spectral density (NPSD) for speech enhancement, which will be herein referred to as NPSD-AAP. In the proposed spectral estimator, both the raw periodogram and the NPSD are smoothed over frequency to reduce their variances if the NPSD has a relatively flat spectrum. Otherwise no smoothing is performed so as to satisfy the high-frequency resolution demand for the non-flat spectrum of the NPSD. The NPSD-AAP provides a low-variance and adaptive-bandwidth estimate of the power spectral density, which could be applied to any frequency-domain speech enhancement algorithms. Especially, the NPSD-AAP is applied to spectral subtraction to suppress the *musical noise* without introducing audible speech distortion. Experimental results confirm the validity of the proposed algorithm.

1. INTRODUCTION

The fast Fourier transform (FFT) is often used to compute the raw periodogram for speech enhancement due to its computational efficiency [1]. However, it is well-known that the raw periodogram is not a good spectral estimate for its large variance, which leads to the *musical noise* problem for most frequency domain speech enhancement algorithms [2]-[11].

There are two ways to suppress the *musical noise*. One is to reduce the variance of the raw periodogram directly, which will be herein referred to as the direct approach. The other is based on other mechanisms, which will be herein referred to as the indirect approach. A typical indirect approach is the *decision-directed* approach of Ephraim and Malah [3] and its improved variants [4]-[8], where the key mechanism is the smoothness of the *priori* signal-to-noise-ratio (SNR) [11]. Another typical indirect approach is to use the psychoacoustic models to suppress the *musical noise*. For example, Virag [12] proposed an oversubtraction scheme based on psychoacoustic models, and Gustafsson *et al.* [13] developed a psychoacoustically motivated audio enhancement algorithm.

Several direct approaches also have been proposed recently. Boll [2] used the magnitude averaging technique to reduce the *musical noise*. It is an effective way to reduce the variance of the raw periodogram by averaging the magnitude over time, but only a limited time averaging is allowed due to the non-stationary characteristic of the speech. Hendriks *et al.* [10] proposed an adaptive time segmentation algorithm to find neighboring wide sense stationary segments for each frame. Hu and Loizou [11] proposed a speech enhancement algorithm based on wavelet thresholding the multitaper spectrum (WTMS), where the WTMS is a low-variance spectral estimator. Gustafsson *et al.* [14] suggested using the Bartlett averaging periodogram (BAP). Both the WTMS and the BAP have a constant time-frequency bandwidth product. Spectral subtraction (SS) with adaptive bandwidths was proposed by Gulzow *et al.* [15], where the bandwidth was adaptively determined by the voice activity detector (VAD). An efficient realization without VAD had been proposed in [15], but it needed to estimate the subband SNR to determine the bandwidth. Obviously, when the narrowband noise is much larger than the speech, the WTMS, the BAP, and the adaptive bandwidths-based SS algorithms may cause audible speech distortion due to their wide main-lobe characteristic.

It is well-known that the SS algorithm often requires an accurate estimate of the noise power spectral density (NPSD). That is to say, the performance of the SS algorithm is somewhat influenced by the accuracy of the NPSD [16]. Whereas, the characteristic of the NPSD is rarely used in most conventional SS algorithms [17]. Considering the harmonic-plus-noise model of speech, the spectral estimator must have high-frequency resolution in order to distinguish two neighboring harmonic spectra. This should be the main reason why the raw periodogram is often used to provide a high-frequency resolution spectral estimator for most of the conventional SS algorithms. In fact, using the high or low frequency resolution spectral estimator dose not make any conspicuous difference for the speech signal. This is because using low-frequency resolution spectral estimator may not cause serious speech distortion and only reduces the amount of the noise reduction for voice at low frequencies, where the residual noise could be masked by the voice in most cases. However, using the high or low frequency resolution spectral estimator for the noise has obvious influences on speech enhancement. For the narrowband noise, using low-frequency resolution spectral estimator may cause audible speech distortion and also may decrease the amount of the noise reduction due to the wide main lobe. Therefore, the structure of the NPSD should determine whether the high-frequency resolution is used or not.

In this paper, we propose to smooth the raw periodogram over frequency adaptively, and basing on the structure of the NPSD, to provide a low-variance and adaptive-bandwidth spectral estimator for speech enhancement, which will be referred to as the structure of NPSD-based adaptive averaging periodogram (NPSD-AAP). The basis of the NPSD-AAP is the flatness of the estimated NPSD: if the estimated NPSD has a relative flat spectrum, adjacent averaging smoothing is applied to both the raw periodogram and the estimated NPSD to significantly reduce their variances; otherwise, no smoothing technique is used to satisfy the high-frequency resolution demand. To validate the proposed spectral estimator, the NPSD-AAP is applied to the SS algorithm to suppress the

 ISSN 2076-1465

musical noise while achieving low levels of speech distortion. Experimental results verify the better performance of the proposed NPSD-AAP-based SS algorithm.

This paper is organized as follows: In section 2, the theory and the implementation of the NPSD-AAP are presented. In section 3, the NPSD-AAP is applied to the SS algorithm. Performance evaluation and conclusions are presented in section 4 and section 5, respectively.

2. NPSD-AAP

2.1 Theory

Let $P_D(f)$ be the NPSD at frequency f. The first derivative of the natural logarithm of the NPSD is selected to measure the flatness, which is defined as

$$\Re(f) = \frac{\partial[\ln E\{P_D(f)\}]}{\partial f} = \frac{1}{E\{P_D(f)\}} \frac{\partial[E\{P_D(f)\}]}{\partial f}, \quad (1)$$

where $\ln\{\bullet\}$ is the natural logarithm function; and $E\{\bullet\}$ is the expectation function. Obviously, the gradient $\Re(f)$ measures the variation of the NPSD. When $\Re(f)$ becomes infinite, the NPSD is increasing or decreasing rapidly at frequency f; On the contrary, if $\Re(f)$ is close to zero, the NPSD is flat. In practice, the discrete Fourier transform (DFT) is used, then (1) can be rewritten as

$$\Re(k,i) = \frac{|E\{P_D(k+i)\} - E\{P_D(k)\}|}{i \cdot E\{P_D(k)\}}, \quad (2)$$

where k is the frequency index, $i = \pm 1, \pm 2, \cdots \pm K_f$, and K_f is the number of the adjacent frequency bins. (2) shows that if and if only $E\{P_D(k)\} = E\{P_D(k+i)\}$,with $i = \pm 1, \pm 2, \cdots \pm K_f$, holds, then $\Re(k,i) \equiv 0$, where the special case is only valid for flat spectra, such as white noise. To be mentioned, (2) is somewhat different from (1) since the absolute operation is applied in (2). $\Re(k,i)$ is still an effective measurement because it has the same physical meaning as $\Re(f)$. (2) can be rewritten as

$$\begin{aligned} &\lambda(k,i) = \frac{E\{P_D(k)\}}{E\{P_D(k+i)\}} \\ &= \begin{cases} \frac{1}{i \cdot \Re(k,i)+1} & E\{P_D(k+i)\} \geq E\{P_D(k)\} \\ \frac{1}{1-i \cdot \Re(k,i)} & \text{otherwise} \end{cases} \end{aligned} \quad (3)$$

where $\lambda(k,i)$ indicates the ratio between the expected value of the NPSD at bin k and that of the NPSD at bin $k+i$. When $\lambda(k,i)$ is close to one, the NPSD at bin k and the NPSD at bin $k+i$ can be averaged because they have the same expected values.

Given the following two hypotheses,

$$\begin{aligned} H_0(k,i) &: \lambda(k,i) = E\{P_D(k)\}/E\{P_D(k+i)\} = 1 \\ H_1(k,i) &: \lambda(k,i) = E\{P_D(k)\}/E\{P_D(k+i)\} \neq 1 \end{aligned} \quad (4)$$

where $i = \pm 1, \pm 2 \cdots \pm K_f$. $H_0(k,i)$ indicates a hypothesis that the NPSD at bins k and $k+i$ have the same expected values; and $H_1(k,i)$ indicates an alternative hypothesis that the NPSD at bins k and $k+i$ have the different expected values. If $H_0(k,i)$ is true, the smooth operation can be applied to bins k and $k+i$ to reduce the variance because they have the same expected values.

2.2 The NPSD estimation and the implementation of the NPSD-AAP

2.2.1 The NPSD estimation

There are two ways to estimate the NPSD. One is based on the VAD, where the NPSD is estimated from the noise-only segment [3],[14],[15]. The other is based on the minimum statistics (MS) approach [16]. Martin has proved that *the variance of the minimum statistics power estimate is smaller than the variance of a single recursively smoothed power estimate* [16]. Thus, the MS approach is applied to estimate the NPSD due to its low-variance characteristic.

We assume that the noise is $d(n)$ and the clean speech is $s(n)$, then the noisy speech $y(n)$ is given by

$$y(n) = s(n) + d(n). \quad (5)$$

We further assume the raw periodogram of $y(n)$, computed by the N-point FFT with Hanning window, is $I_Y(k,l)$, where l is the frame index, and $k = 0, 1 \cdots N-1$. The NPSD can be estimated by the following two steps:

1) Recursive smoothing of the raw periodgoram $I_Y(k,l)$ leads to

$$P_Y(k,l) = \alpha P_Y(k-1,l) + (1-\alpha) I_Y(k,l), \quad (6)$$

where α is a forgetting rate parameter.

2) The minimum from the Ω consecutive samples of $P_Y(k,l)$ is the estimated NPSD, which is given by

$$P_D(k,l) = \beta_c \min(P_Y(k,l) | l-\Omega+1, \cdots l), \quad (7)$$

where β_c compensates the bias. In the rest of this paper, the frame index l is discarded without causing confusion.

2.2.2 The NPSD-AAP

The NPSD estimated by the MS approach has a low-variance characteristic, so we can define

$$\hat{\lambda}(k,i) = P_D(k)/P_D(k+i), \quad (8)$$

then the decision rule for the binary hypothesis problem defined in (4) can be given by

$$\begin{cases} H_0(k,i) \text{ true}, w(k,i) = 1, if \hat{\lambda}(k,i) \in \left[\frac{1}{\lambda_{\text{th}}} \quad \lambda_{\text{th}}\right] \\ H_1(k,i) \text{ true}, w(k,i) = 0, if \hat{\lambda}(k,i) \notin \left[\frac{1}{\lambda_{\text{th}}} \quad \lambda_{\text{th}}\right] \end{cases} \quad (9)$$

where $w(k,i) = 1$ indicates accepting $H_0(k,i)$; otherwise, $w(k,i)$ is set to zero. $\lambda_{\text{th}} \geq 1$ is the threshold for accepting or rejecting $H_0(k,i)$, which can be obtained by the false alarm rate (FAR), where the FAR is the probability of choosing $H_1(k,i)$ when in fact $H_0(k,l)$ is true. Obviously, if $H_0(k,i)$ is accepted, the hypothesis $H_0(i,k)$ also must be accepted for symmetry. This is the reason why the interval is $\left[\ 1/\lambda_{\text{th}} \quad \lambda_{\text{th}}\ \right]$ in (9). The FAR is given by

$$P_{fa} = \int_0^{1/\lambda_{\text{th}}} f_{\hat{\lambda}(k,i)|\lambda(k,i)=1}(x)dx + \int_{\lambda_{\text{th}}}^{\infty} f_{\hat{\lambda}(k,i)|\lambda(k,i)=1}(x)dx \quad (10)$$

where $f_{\hat{\lambda}(k,i)|\lambda(k,i)=1}(x)$ is the conditional probability density function (pdf) of $\hat{\lambda}(k,i)$ given $\lambda(k,i) = 1$. The threshold

For all time frame l

1) Estimate the NPSD by the MS approach at all frequency bins k: $P_D(k)$.
2) Calculate the ratio between the NPSD at bin k and the NPSD at bin $k+i$ $\hat{\lambda}(k,i)$ using Eq. (8)
3) Determine the weighted factor $w(k,i)$ using Eq. (9).
4) Smooth the raw periodogram $I_Y(k)$ and the estimate NPSD $P_D(k)$ using Eqs. (11) and (12).
5) Compute the gain function $G(k)$ using Eq. (13).
6) Calculate the estimate of the clean speech $\hat{s}(n)$ using Eq. (14).

Figure 1: NPSD-AAP-SS algorithm.

Table 1: Parameter values for the NPSD-AAP-SS, where N is the frame length, and M is the frame shift parameter.

$f_s = 16$kHz	$\alpha = 0.8$	$\beta_c = 1.85$	$\Omega = 120$	$K_f = 6$
$\lambda_{\text{th}} = 3$	$\beta = 3$	$G_{\min} = -20$dB	$N = 512$	$M = N/2$

λ_{th} relies on the window type, the overlapping parameter, the forgetting rate parameter α, and the parameter Ω used in the MS approach. In this paper, the threshold λ_{th} is obtained by Monte Carlo method and a typical value is 3.

Based on (8) and (9), the smoothed periodogram and the smoothed NPSD at bin k can be computed by

$$\tilde{I}_Y(k) = \sum_{i=-K_f}^{K_f} w(k,i) I_Y(k+i) \Bigg/ \sum_{i=-K_f}^{K_f} w(k,i), \qquad (11)$$

$$\tilde{P}_D(k) = \sum_{i=-K_f}^{K_f} w(k,i) P_D(k+i) \Bigg/ \sum_{i=-K_f}^{K_f} w(k,i). \qquad (12)$$

Implementing (8) directly requires $K_f/2+1$ divisions for each bin as considering the symmetry of the raw periodogram. When K_f is not too large, the computation load of the NPSD-AAP does not increase too much compared with that of the raw periodogram.

3. SPECTRAL SUBTRACTION BASED ON THE NPSD-AAP

The gain function of the SS algorithm based on the NPSD-AAP (NPSD-AAP-SS) is given by

$$G(k) = \max\left\{ \frac{\max\{\bar{I}_Y(k) - \beta\bar{P}_D(k), 0\}}{\max\{\bar{I}_Y(k) - \beta\bar{P}_D(k), 0\} + \bar{P}_D(k)}, G_{\min} \right\}, \qquad (13)$$

where $\beta > 1$ is the oversubtraction factor, and $G_{\min}$ is the minimum gain value. After the gain function is obtained, the enhanced speech could be computed by

$$\hat{s}(n) = \texttt{IFFT}\{G(k)Y(k)\}, \qquad (14)$$

where $Y(k)$ is the FFT of the noisy speech $y(n)$ and $\hat{s}(n)$ is the estimate of the clean speech $s(n)$.

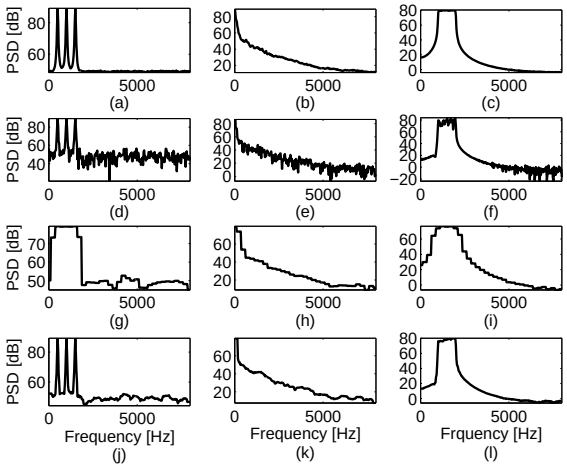

Figure 2: Comparison of the spectral estimators. (a)(b)(c) in the first row show the sinusoids in WGN, car noise, and bandlimited noise, respectively. (d)(e)(f) in the second row and (g)(h)(i) in the third row show the raw periodogram and the BAP, respectively. (j)(k)(l) in the last row illustrate the NPSD-AAP.

Table 2: The RIMSE of the three spectral estimators, including the raw periodogram, the BAP, and the NPSD-AAP.

Noise/Algorithms	the raw periodogram	the BAP	the NPSD-AAP
Sinusoids in WGN	0.44	27.33	0.42
Bandlimited WGN	10.12	4.69	3.76
WGN	256.35	93.90	89.09

We summarize the implementation of the NPSD-AAP-SS algorithm in Figure 1. All of the signals used in this paper are sampled at $f_s = 16$kHz, and the parameter values for the proposed algorithm are shown in Table 1.

4. PERFORMANCE EVALUATION

In this section, we first compare the NPSD-AAP with the raw periodogram and the BAP to show the low-variance and adaptive-bandwidth characteristics of the proposed spectral estimator by using three types of noise, including sinusoids buried in white Gaussian noise (WGN), car noise, and bandlimited WGN. The frame length for the raw periodogram and the BAP is 512 and 64, respectively; and [512/64] subblocks are averaged to obtain the BAP in the simulation. Second, the NPSD-AAP-SS is compared with the raw periodgoram-based SS (RP-SS) and the BAP-based SS (BAP-SS) algorithms.

4.1 Comparison of the spectral estimators

Three types of noise, including sinusoids buried in WGN, car noise, and bandlimited WGN, are shown in the first row of Figure 2, respectively. The raw periodogram and the BAP of these noise signals are shown in the second and the third rows of Figure 2, respectively. The last row of Figure 2 depicts the estimation results of the NPSD-AAP. In the proposed NPSD-AAP algorithm, the variance is better reduced for the flat spectrum of the NPSD, and the frequency resolution is high enough for the non-flat spectrum of the NPSD. Whereas, the raw periodogram has a large variance with high-frequency resolution; and the BAP has a low variance

Table 3: Comparison of the log-spectral distance (LSD) for the three SS algorithms, including the RP-SS, the BAP-SS, and the proposed NPSD-AAP-SS algorithms.

LSD [dB]												
Algorithm	RP-SS				BAP-SS				NPSD-AAP-SS			
Input Segmental SNR	-5	0	5	10	-5	0	5	10	-5	0	5	10
WGN	8.20	4.87	2.94	1.81	5.65	3.51	2.55	1.78	**5.61**	**3.42**	**2.49**	**1.76**
Bandlimited WGN	4.81	3.15	2.06	1.38	4.28	2.81	1.81	1.27	**3.97**	**2.58**	**1.70**	**1.26**
Sinusoids in WGN	**3.34**	**2.02**	**1.27**	**0.86**	5.19	3.01	2.18	1.38	3.44	2.08	1.34	0.95
Car noise	4.28	1.84	1.20	0.89	3.95	2.12	1.38	1.16	**2.51**	**1.76**	**1.07**	**0.78**
Babble	6.26	3.92	2.47	1.60	**5.36**	**3.31**	2.09	1.41	5.47	3.38	**2.07**	**1.36**

with low-frequency resolution. The proposed NPSD-AAP is a low-variance and adaptive-bandwidth spectral estimator, which could make a good trade-off between the variance and the frequency resolution based on the flatness of the NPSD.

To give a quantitative result, the root integrated mean square error (RIMSE) defined in [18] is used to evaluate the performance of the three spectral estimators, where the results are shown in Table 2. The best performance of the NPSD-AAP is further confirmed by the minimum RIMSE among the three spectral estimators for any type of noise. The RIMSE of the car noise is not presented in Table 2 for which the car noise is non-stationary, and it is unable to define its expected value. Even for the WGN, the variance is only reduced by a factor of less than $2K_f + 1$. The main reason is that the Hanning window and the overlap used in the raw periodogram reduce the independence.

4.2 Comparison of the three SS algorithms

The NPSD-AAP-SS is compared with the RP-SS and the BAP-SS, where the noise signals include two types of artificial noise as shown in Figure 2(a) and (c), and three noise signals (WGN, car noise, and babble) taken from the Noisex92 database [19]. More than 400 clean speech samples are taken from the TIMIT database [20]. These clean speech samples are summed up to about 20 minutes without intervening pauses and degraded by the various noise types with segmental SNRs in the range [-5 10]dB. The log-spectral distance (LSD) [21], the perceptual evaluation of speech quality (PESQ) [22], and the speech spectrograms are used to give the objective comparison results. The results of the LSD are shown in Table 3.

As shown in Table 3, the proposed NPSD-AAP-SS algorithm has a smaller LSD than the other two SS algorithms for most cases. For the broadband noise with small dynamic range, such as the WGN and the bandlimited WGN, the performance of the BAP-SS is comparable with that of the NPSD-AAP-SS. The main reason is that the BAP and the NPSD-AAP nearly have the same frequency resolution, and the NPSD-AAP is also a low-frequency resolution and low-variance spectral estimator just like the BAP; while the performance of the RP-SS is the worst due to the large variance of the raw periodogram. For the narrowband noise with large dynamic range, such as the sinusoids buried in WGN, the BAP-SS is the worst due to the low-frequency resolution of the BAP; the proposed NPAD-AAP-SS has nearly the same performance with the RP-SS because the NPSD-AAP could reduce the variance and achieve high-frequency resolution simultaneously based on the flatness of the NPSD. For the non-stationary broadband noise with medium dynamic range, such as the car noise and the babble, the best performance of the proposed algorithm reveals that the NPSD-AAP-SS is not seriously deteriorated when the MS approach underestimates the NPSD.

Table 4: Comparison of the PESQ improvement for the three SS algorithms at an input segmental SNR = 0dB.

Noise	PESQ Improvement		
Algorithm	RP-SS	BAP-SS	NPSD-AAP-SS
WGN	0.59	**0.62**	**0.62**
Bandlimited WGN	0.60	0.57	**0.72**
Sinusoids in WGN	**0.59**	0.09	0.50
Car noise	**0.50**	0.24	**0.50**
Babble noise	0.08	**0.13**	**0.13**

The PESQ has been found to have a good correlation overall with the mean opinion score (MOS), so it is used to further confirm the better performance of the proposed NPSD-AAP-SS, where the PESQ improvement for the three SS algorithms is shown in Table 4. The NPSD-AAP-SS has a higher PESQ improvement for most cases, which is consistent with the results of the LSD.

To give indications of the structure of the residual noise and the speech distortion, an example of spectrograms of the noisy and the enhanced speech samples is presented in Figure 3, where the clean speech is corrupted by the sinusoids buried in WGN at -5dB. For the proposed algorithm, the *musical noise* is reduced without introducing more speech distortion. The BAP-SS suppresses the *musical noise* at the expense of more speech distortion at the sinusoidal frequencies, while the RP-SS still has annoying *musical noise*.

Informal listening tests further show that the RP-SS still has annoying *musical noise* for the five types of noise, while the BAP-SS has audible speech distortion for three types of noise including the sinusoids buried in WGN, the bandlimited WGN, and the car noise. Both the BAP-SS and the NPSD-AAP-SS can effectively suppress the *musical noise*, while the NPSD-AAP-SS does not introduce audible speech distortion for the five types of noise.

5. CONCLUSIONS

This paper proposes an adaptive averaging periodogram based on the structure of noise power spectral density, where the proposed NPSD-AAP could be applied to any frequency-domain speech enhancement algorithms that need periodogram estimation. Compared with the raw periodogram and the BAP, the NPSD-AAP provides a low-variance and adaptive-bandwidth estimate of the power spectral density, which could achieve high-frequency resolution of the NPSD

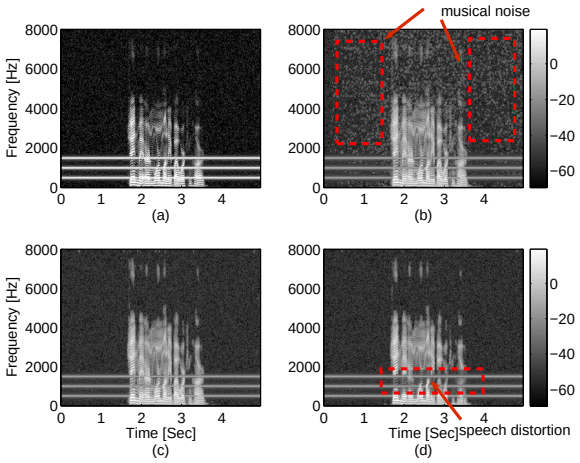

Figure 3: Speech spectrograms for the clean speech corrupted by sinusoids buried in WGN at -5dB. (a) Noisy speech. (b) Enhanced speech by the RP-SS. (c) Enhanced speech by the NPSD-AAP-SS. (d) Enhanced speech by the BAP-SS.

and reduce the variance of the raw periodogram simultaneously based on the flatness of the NPSD. Compared with the raw periodogram-based and the BAP-based spectral subtraction algorithms, the NPSD-AAP-based SS algorithm could suppress the *musical noise* without causing more speech distortion for any type of noise.

We wish to emphasize that there are at least two ways to improve the performance of the NPSD-AAP-SS. One is to select the oversubtraction parameter β in (13) according to the speech presence probability (SPP) [23]. When the SPP is close to one, β should be small to reduce speech distortion; otherwise, if the SPP is close to zero, β should be large to suppress the noise. The other way is to further reduce the nonstationary noise by cepstral smoothing technique [7].

REFERENCES

[1] J. Benesty, S. Makino and J. Chen, *Speech Enhancement*. Berlin: Springer-Verlag, 2005.

[2] S. F. Boll, Suppression of acoustic noise in speech using spectral subtraction, IEEE Trans. on Acoustics, Speech, and Signal Process., 27(1979) 113-120.

[3] Y. Ephraim and D. Malah, Speech enhancement using a minimum mean square error short-time spectral amplitude estimator, IEEE Trans. Acoustics, Speech, and Signal Process., 32(1984) 1109-1121.

[4] I. Cohen, Relaxed statistical model for speech enhancement and *a priori* SNR estimation, IEEE Trans. Speech and Audio Process., 13(2005) 870-881.

[5] O. Cappe, Elimination of musical noise phenomenon with Ephraim and Malah noise suppressor, IEEE Trans. Speech and Audio Process., 2(1994) 345-349.

[6] C. Plapous, C. Marro and P. Scalart, Improved signal-to-noise ratio estimation for speech enhancement, IEEE Trans. Audio, Speech, and Lang. process., 14(2006) 2098-2108.

[7] C. Breithaupt, T. Gerkmann and R. Martin, Cepstral smoothing of spectral filter gains for speech enhancement without musical noise, IEEE Signal Process. Letter, 14(2007) 1036-1039.

[8] T. Esch and P. Vary, Efficient musical noise suppression for speech enhancement systems, in Proc. IEEE ICASSP. Taipei, Taiwan, April, 2009, pp. 4409-4412.

[9] C. Zheng, X. Li, J. Chen and J. Tian, Speech enhancement based on adaptive averaging periodogram, ACTA ACUSTICA, 32(2007) 461-467. (*in Chinese*)

[10] R. C. Hendriks, R. Heusdens and J. Jensen, Adaptive time segmentation for improved speech enhancement, IEEE Transactions on Audio, Speech, and Language Processing, 14(2006) 2064-2074.

[11] Y. Hu and P. C. Loizou, Speech enhancement based on wavelet thresholding the multitaper spectrum, IEEE Trans. on Speech and Audio Process., 12(2004) 59-67.

[12] N. Virag, Single channel speech enhancement based on masking properties of the human auditory system, IEEE Trans. on Speech and Audio Process., 7(1999) 126-137.

[13] S. Gustafsson, P. Jax and P. Vary, A novel psychoacoustically motivated audio enhancement algorithm preserving background noise characteristics, in Proc. IEEE ICASSP. Seattle, USA, May 1998, pp. 397-400.

[14] H. Gustafsson, S. E. Nordholm and I. Claesson, Spectral subtraction using reduced delay convolution and adaptive averaging, IEEE Trans. on Speech and Audio Process., 9(2001) 799-807.

[15] T. Gulzow, T. Ludwig, U. Heute, Spectral-subtraction speech enhancement in multirate systems with and without non-uniform and adaptive bandwidths. Signal Process., 83(2003) 1613-1631.

[16] R. Martin, Bias compensation methods for minimum statistics noise power spectral density estimation, Signal Processing, 86(2006) 1215-1229.

[17] K. Manohar and P. Rao, Speech enhancement in nonstationary noise enviornments using noise properties, Speech Communication, 48(2006) 96-109.

[18] P. Stoica and N. Sandgren, Smoothed nonparametric spectral estimation via cepstrum thresholding, IEEE Signal Process. Mag., 23(2006) 34-45.

[19] A. Varga and H. J. M. Steeneken, Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems, Speech Commun., 12(1993) 247-251.

[20] J. S. Garofolo, DARPA TIMIT acoustic-phonetic speech database, Nat. Inst. Standards Technol. (NIST), 1988.

[21] S. R. Quakenbush, T. P. Barnwell, and M. A. Clements, *Objective measures of speech quality*. Englewood Cliffs, NJ: Prentice-Hall, 1988.

[22] Recommendation P.862: Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs, ITU-T, 2001.

[23] T. Gerkmann, C. Breithaupt and R. Martin, Improved *A Posteriori* Speech Presence Probability Estimation Based on a Likelihood Ratio With Fixed Priors, IEEE Trans. on Audio, Speech, and Lang. Process., 16(2008) 910-919.