Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2010, Article ID 395048, 11 pages
doi:10.1155/2010/395048

Research Article

Improved Noise Minimum Statistics Estimation Algorithm for Using in a Speech-Passing Noise-Rejecting Headset

Saeed Seyedtabaee and Hamze Moazami Goodarzi

Department of Electrical Engineering, Engineering Faculty, Shahed University, P.O. Box 18155/159, Tehran, Iran

Correspondence should be addressed to Saeed Seyedtabaee, gstabaii@gmail.com

Received 23 August 2009; Revised 7 March 2010; Accepted 8 May 2010

Academic Editor: Igor Djurović

This paper deals with configuration of an algorithm to be used in a speech-passing angle grinder noise-canceling headset. Angle grinder noise is annoying and interrupts ordinary oral communication. Meaning that, low SNR noisy condition is ahead. Since variation in angle grinder working condition changes noise statistics, the noise will be nonstationary with possible jumps in its power. Studies are conducted for picking an appropriate algorithm. A modified version of the well-known spectral subtraction shows superior performance against alternate methods. Noise estimation is calculated through a multi-band fast adapting scheme. The algorithm is adapted very quickly to the non-stationary noise environment while inflecting minimum musical noise and speech distortion on the processed signal. Objective and subjective measures illustrating the performance of the proposed method are introduced.

1. Introduction

Industrial site noises jeopardize workers health condition. To alleviate the risk, a passive protecting headset may be worn. It gives good attenuation of ambient noise in the upper frequency band and some how medium protection in below 500 Hz. Along with the noise, the oral communication link is also disrupted that should not be.

To improve the working condition, a type of active headset is designed that allows receiving speech while its capacity in reducing noise is still in place. The headset in its simplest form consists of a microphone, a battery-powered processing unit, and one speaker in one of the ear cups (or separate sets of microphone, processing unit, and speaker, one for each ear cup) as shown in Figure 1.

Microphone may receive noise, speech, or noisy speech signal. The processing unit is expected to enhance the speech signal and to reduce the noise in any case.

Speech enhancement is one of the most important topics in signal processing. Enhancement techniques can be classified into single and multichannel classes. Single-channel systems are the most common real-time scenario algorithms, since the second channel is not available in most of the applications, for example mobile communication, hearing aids, speech recognition systems, and the case of speech-passing noise-canceling headset. The single-channel systems are easy to build and comparatively less expensive than the multiple input systems. Nevertheless, they constitute one of the most difficult situations of speech enhancement, since no reference signal is available, and clean speech cannot be statistically preprocessed prior to getting affected by noise.

Wide variety of algorithms has been developed for single microphone speech enhancement. In *waveform filtering* class, only limited assumptions are made about the specific nature of the underlying signal. The most prominent examples of waveform processing are the spectral subtraction method [1], spectral or cepstral restoration [2], Wiener filter [3], the Wiener filtering extensions [4, 5], and adaptive filtering type [6].

Other examples include schemes that employ wavelets [7], modifications of the iterative Wiener filter and the Kalman filter [8, 9]. Perceptual Kalman filtering for speech enhancement in [10, 11] and Rao-Blackwellized particle filtering (RBPF) in [12] are elaborated.

Nondiagonal time-frequency estimators that introduce less musical noise backing up with an adaptive audio block threshold setting algorithm have been studied in [13].

In stochastic model-based denoising methods, a stochastic parametric model for a speech signal is used instead of a general waveform model. One statistical model method is discussed in [14]. Accurate modeling and estimation of speech and noise via Hidden Markov Models are proposed in [15]. A minimum mean square error approach for denoising that relies on a combined stochastic and deterministic speech model is discussed in [16]. Formant tracking linear prediction (LP) model for noisy speech processing is reported in [17].

Among all this wide range of methods, the spectral subtraction-based algorithm is known for its (1) simplicity in implementation, (2) high power in eliminating noise, and (3) high speed. The most important problems with spectral subtraction are speech distortion and residual noise that is called musical noise. These problems are due to nonaccurate noise estimation in each frame and differences between the estimated clean and original signal.

A very challenging task of spectral subtraction speech enhancement algorithms is noise spectrum estimation. Originally, it requires the silent period to be detected. An algorithm that does not require explicit speech/pause detection and can update noise estimate even from noisy speech sections is proposed in [18]. The algorithm is based on finding the minimum statistics of noisy speech for each subband over a time window. Its major drawback is that when the noise floor jumps, it takes slightly more one window length to update the noise spectrum estimate. Updating continuously the noise estimate is suggested in [19]. However, the algorithm cannot distinguish between a rise in noise power and a rise in speech power. In the algorithm, there is a very sophisticated formula for computing gain factors for each subband. The gain factors overestimate the noise and permit gradual suppression of certain subbands as their speech contribution decreases. Hirsch and Ehrlicher [20] produce subband energy histograms from past spectral values below the adaptation threshold over a duration window and choose the maximum noise level to update the noise estimate. The major drawback of their method is that it fails to update the noise estimate when the noise floor increases abruptly and stays at that level. The method proposed in [21] uses a recursive equation to smooth and update noise power estimate with a smoothing parameter related to a priori SNR. This method needs more time to estimate the noise, especially when the noise floor jumps. The drawback of the algorithm in [22] is its large latency. Some improved algorithms have been proposed in [23–25]. These also suffer from the similar problem. The authors in [26] propose an algorithm based on temporal quantile and make use of the fact that even within speech sections of input signal, not all frequency bands are permanently occupied with speech. Rather, for a significant percentage of time the energy within each frequency band equals the noise level. This method suffers from computational complexity and requires higher memory and therefore is not really recommended for real-time systems.

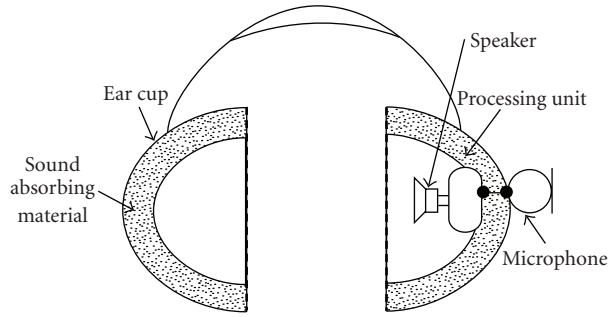

Figure 1: The proposed headset.

A method that most fits our speech-passing noise-rejecting headset design is the one that (1) renders acceptable results, (2) has low computational cost, and (3) enjoys simplicity in implementation. Our primary goal is the design of a headset that combats the angle grinder noise. Of course, it can be easily extended to the other rotating devices noise.

From this point of view, the adaptive notch filter method was thoroughly investigated. Even though, the case is similar to the problem discussed in [6]; in this case, the application of various types of adaptive notch filter remained fruitless.

The improvement of spectral subtraction was the next attempt [27]. Improved spectral subtraction method appeared strong in forming effective algorithm for rejection noise. The algorithm embodies fast adapting capability, as sharp change in angle grinder noise characteristics is noticed. Using subwindows makes noise estimate updating faster and enables tracking jumps in the noise power. Another point is that *a priori* qualitative coarse knowledge of the spectrum of the angle grinder noise is easily available that can be incorporated into the algorithm. This led us to the proposed combined multiband fast adapting spectral subtraction method. Angle grinder noise spectrum is not flat, so multiband noise minimum statistics estimation is implemented. This is inevitably required for the developing of an algorithm that takes the musical noise and speech distortion under control.

This paper reports our latest achievements. In Section 2, we analyze angle grinder noise. The adaptive notch is discussed in Section 3. The spectral subtraction is reviewed in Section 4. In Section 5, our noise estimation algorithm is disclosed. Performance evaluation is presented in Section 6. Section 7 contains the experimental set-up and the test results. Finally, conclusion in Section 8 ends up this discussion.

2. Angle Grinder Noise Analysis

Angle grinder acoustic noise specs change as the device engagement condition with a part varies. The characteristics of the noise also depend on the brand and size of angle grinder. The material of the engaged part also contributes to the generated sound, as each part generates sound of its own. Figure 2 shows the noise waveform of a typical angle grinder.

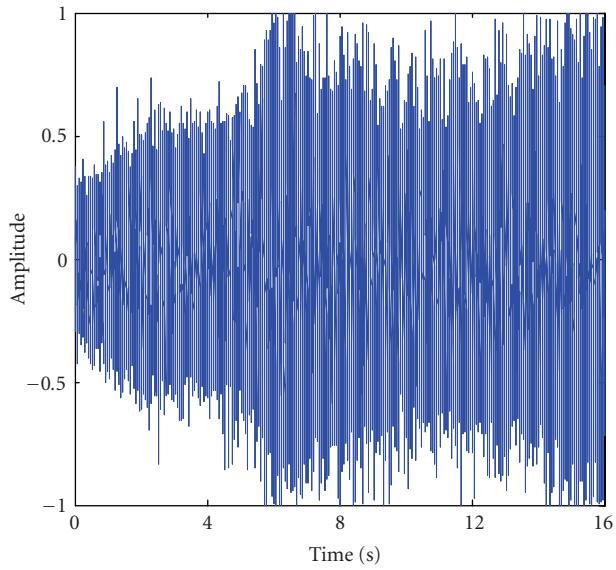

Figure 2: Waveform of a typical angle grinder noise.

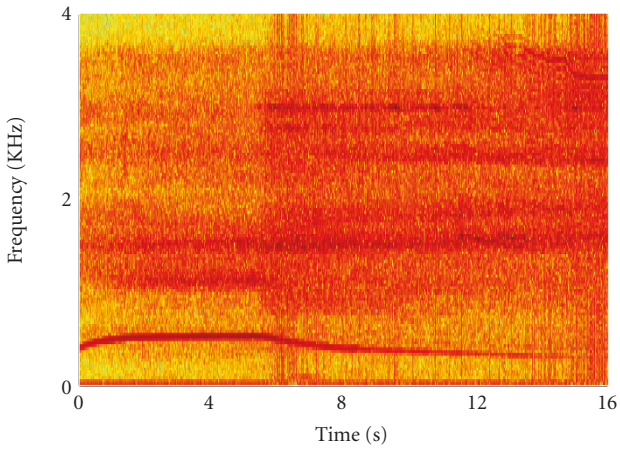

Figure 3: Angle grinder noise spectrograph.

Spectral content of the angle grinder noise is an important factor to be considered in the development of our noise removal system. The noise spectrum is typically comprised of a wide-band section and some peaks that have been referred to as a periodic part plus its harmonics. Figure 3 shows the spectrograph of the angle grinder noise. Dark lines indicate existence of strong frequency components in the spectrum. The frequency is related to the rotation speed of the angle grinder.

It also reveals that the noise is wide band and each frequency bin contains some of the noise power. The noise spectrum is not flat. Variation in noise spectrum due to change in working condition is apparent from Figure 3. Major frequency components of the noise change in both amplitude and frequency. Generation of new frequency components is apparent from the spectrograph. Change in noise spectrum means that we are facing a type of nonstationary behavior.

3. Adaptive Notch Filter Method

From the analysis of angle grinder noise, it is discovered that some of the energy is concentrated in specific frequency components and their harmonics. In line with this type of analysis, we use adaptive notch algorithm discussed in [6]. The algorithm is adaptive and is able to track change in frequency variations. The system employs a cascade of three second-order adaptive notch/band-pass filters based on Gray-Markel lattice structure. This structure ensures the high stability of the adaptive system. A Newton type algorithm is used for updating the filter coefficients that enjoy fast adaptation. In addition, a new algorithm using adaptive filtering with averaging (AFA) is also verified. The main advantages of AFA algorithm could be summarized as follows: high convergence rate comparable to that of the recursive least squares (RLSs) algorithm and at the same time low computational-complexity.

Adaptive noise-canceling systems are often two channel types, in which one channel is dedicated to the noisy signal and the other captures the reference signal. In modification to the adaptive systems, when *a priori* knowledge of the noise fundamental frequency exists, coarse value of the fundamental frequency is introduced to the algorithm; this obviates further need to the reference signal, and a single microphone adaptive system gets applicable.

4. Spectral Subtraction Method

The main assumption in the spectral subtraction method is that the speech signal is corrupted by an uncorrelated additive noise. This is a true assumption in the most real-world cases. A speech signal $s(n)$ that has been degraded by an uncorrelated additive noise signal $\underline{n}(n)$ is written as follows:

$$x(n) = s(n) + \underline{n}(n). \tag{1}$$

The other assumption is that the noise power spectrum in each window W is a slowly varying process; thus it can be assumed stationary in each window. The power spectrum of the noisy signal in window W can be represented by

$$\begin{aligned} |X_w(k)|^2 &= |S_w(k)|^2 + |N_w(k)|^2 \\ &\quad + S_w(k)N_w^*(k) + S_w^*(k)N_w(k), \end{aligned} \tag{2}$$

where $S_w^*(k)$ and $N_w^*(k)$ represent the complex conjugate of $S_w(k)$ and $N_w(k)$, respectively. The functions $|S_w(k)|^2$ and $|N_w(k)|^2$ are referred to as the short-time power spectrum of the speech and noise, respectively. Here, the short-term Fourier transform (STFT) of $X_w(k)$ is obtained by

$$\begin{aligned} X_w(k) &= \sum_{n=0}^{N-1} x(\lambda R + n)W(n)e^{-j2\pi(kn/N)} \\ &= |X_w(\lambda, k)|e^{j\Phi(\lambda,k)}, \end{aligned} \tag{3}$$

where λ, N, and $100 \times (N - R)/N$ are the frame index, the frame length, and the overlapping percentage, respectively. $\Phi(\lambda, k)$ is the phase of the corrupted noisy signal.

In (2), the term $|N_w(k)|^2$, cross-terms $S_w(k)N_w^*(k)$ and $S_w^*(k)N_w(k)$ cannot be obtained directly and are approximated by $E[|N_w(k)|^2]$, $E[S_w(k)N_w^*(k)]$, and $E[S_w^*(k)N_w(k)]$. Where $E[\cdot]$ denotes the expectation operator. If we assume that $\underline{n}(k)$ is zero mean and uncorrelated with $s(k)$, then the cross-terms $E[S_w(k)N_w^*(k)]$ and $E[S_w^*(k)N_w(k)]$ are reduced to zero. Thus, from the above assumptions, the estimate of the clean speech is given by

$$\left|\hat{S}_w(k)\right|^2 = |X_w(k)|^2 - E|N_w(k)|^2. \tag{4}$$

Typically, $E[|N(k)|^2]$ is estimated during the silent periods and denoted by $|\hat{N}(k)|^2$. With respect to the assumption that the noise is stationary in each window, $|\hat{N}(k)|^2$ is regarded as the noise power estimate.

To construct the denoised signal, two steps are undertaken. First, the estimated noise minimum statistics amplitude is reduced from the noisy speech spectrum amplitude. In the second step then, the result is combined with the phase of the noisy speech signal spectrum. The described operations are managed through using an inverse discrete Fourier transform that yields the processed denoised signal as follows

$$\hat{s}_w(n) = IDFT\left[\left|\hat{S}_w(k)\right| e^{j\Phi(k)}\right]. \tag{5}$$

The phase of the noisy signal is not modified since human perception is not sensitive to the phase [28]. However, in a recent work [29], the authors have shown that at lower SNRs, below 0 db, the phase error causes considerable speech distortion.

Since the average magnitude of an instantaneous noise spectrum does not follow truly sharp peaks of the noise, an annoying residual noise, called musical noise, appears after applying spectral subtraction method. Most of the research in the past decade has been focused on the ways to combat the problem of the musical noise. It is literally impossible to minimize musical noise without affecting the speech quality, and hence, there should be a trade-off between the amount of noise reduction and speech distortion.

The proposed method in [30] is one of the earliest methods to reduce residual noise. Modifications that we made to the original spectral subtraction method are (1) subtracting an overestimate of the noise power spectrum and (2) preventing the resultant spectrum from going below a preset minimum level (spectral floor). The proposed algorithm is expressed by

$$\left|\hat{S}_w(k)\right|^2 = \begin{cases} |X_w(k)|^2 - \alpha\left|\hat{N}_w(k)\right|^2 & \text{if } |X_w(k)|^2 \geq \alpha\left|\hat{N}_w(k)\right|^2, \\ \beta\left|\hat{N}_w(k)\right|^2 & \text{else,} \end{cases} \tag{6}$$

where α is the oversubtraction factor, and β is the spectral floor parameter. The oversubtraction factor α depends on the segmental noisy signal to noise ratio (*NSNR*) that is calculated for every frame by:

$$\mathrm{SNR}_i = 10\log\left[\frac{\sum_{k=b_i}^{e_i}|X_i(k)|^2}{\sum_{k=b_i}^{e_i}\left|\hat{N}_i(k)\right|^2}\right], \tag{7}$$

where b_i and e_i are the beginning and ending frequency bins of the i_{th} frequency band. In this definition, it is allowed that the overall frequency band divided into several subbands. The oversubtraction factor α is calculated by

$$\alpha = \begin{cases} 1 & \mathrm{NSNR} \geq 20\,\mathrm{dB}, \\ \alpha_0 - \dfrac{3}{20}\mathrm{NSNR} & -5\mathrm{dB} \leq \mathrm{NSNR} \leq 20\,\mathrm{dB}, \\ 5 & \mathrm{NSNR} \leq -5\,\mathrm{dB}, \end{cases} \tag{8}$$

where $\alpha_0 = 4$ is the desired value of α at 0 db NSNR.

5. Noise Minimum Statistics Estimation: The Proposed Multiband Fast Adaptive Algorithm

5.1. The Initial Algorithm:The Martin's Method. A very challenging task of spectral subtraction speech enhancement algorithms is noise spectrum estimation. For estimating stationary noise specifications, the first 100–200 ms of each noisy signal are usually assumed pure noise and used to estimate the noise for over the time [31]. For estimation of nonstationary noise, the noise spectrum needs to be estimated and updated continuously. To do so, we need a voice activity detector (*VAD*) to find silence frames for updating noise estimation [32]. In a nonstationary noise case or low *SNR* situations, nonspeech/pause section detection reliability is a concern. In [18], the author proposes an algorithm that does not require explicit speech/pause detection and can update noise estimation even from noisy speech sections. The minimum statistics noise tracking method is based on the observation that even during speech activity a short-term power spectral density estimate of the noisy signal frequently decays to values that are representative of the noise power level. Thus, by tracking the minimum power within finite *(D) PSD* frames, large enough to bridge high power speech segments, the noise floor can be estimated [33].

The smoothed power spectrum of noisy speech $P_x(\lambda, k)$ is calculated with a first-order recursive equation as follows:

$$P_x(\lambda, k) = \eta P_x(\lambda - 1, k) + (1 - \eta)|X(\lambda, k)|^2, \tag{9}$$

where λ and k are the frame and the frequency bin indices, respectively. η is a smoothing constant where value is to be set appropriately between zero and one. Often a constant value of 0.85 to 0.95 is suggested [33].

If $x(n)$ can be assumed stationary with a relatively small span of correlation and for a large frame size, the real and imaginary part of the Fourier transform coefficients, $X(\lambda, k)$, can be considered independent and modeled as zero mean Gaussian random variables [34]. Under this assumption,

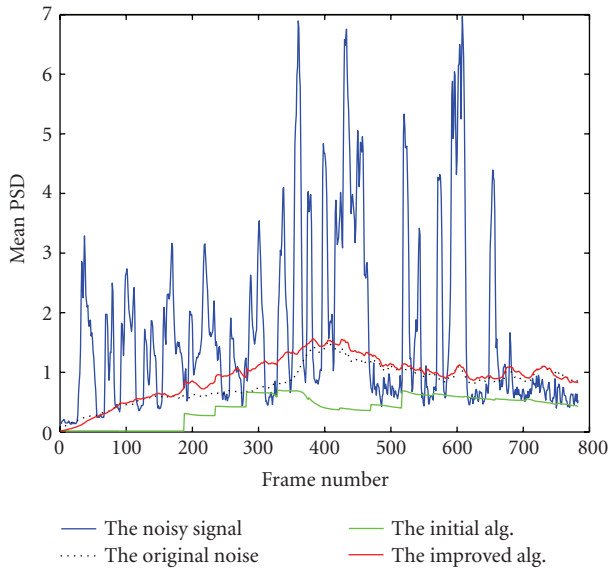

FIGURE 4: The average smoothed PSD of the noisy speech, the noise, the initial method estimate and our algorithm estimate.

each periodogram bin is an exponentially distributed random variable. If the condition holds, an optimal smoothing constant derived in [33] can be employed that enhances the performance

$$\eta_{\text{opt}}(\lambda,k) = \frac{1}{1+\left((P_x(\lambda-1,k))/(\sigma_n^2(\lambda,k))-1\right)^2}, \quad (10)$$

where $\sigma_n^2(\lambda,k)$, the true *PSD* of the noise, can be replaced by its latest estimate, $P_n(\lambda,k)$. More works on this subject have recently been reported in [35]. Dependency of the optimal value of η on λ, k and noise Power Density Frequency (PDF) increases its computation burden while, its allowable range (0.85 to 0.95) is limited, and there is uncertainty about PDF of the (non stationary) noise. This justifies using an average value that is calculated occasionally, instead of the nonoptimal exact value computation in each iteration.

5.2. Noise Spectral Minimum Estimation. Since spectrum of noisy speech signal often decays to the spectrum of noise, we can get an estimate of the noise in a time window of about 0.8–1.4 s. This corresponds to finding the minimum among a number (D) of consecutive *PSD*s, $P_x(\lambda,k)$, as follows:

$$P_{D\min}(\lambda_D,k) = \min\{P_x(\lambda_D-j,k)\},$$
$$j = 0\cdots D-1, \quad \lambda_D = i*L, \quad (11)$$

where i is the estimation iteration number. The calculated spectral minimum, then, is used in the future frames, ($\lambda > \lambda_D$), for spectral subtraction. The equation may be updated in every and each λ step, $L = 1$, then $k\times(D-1)$ compare operations are needed per step. However, if it is computed after every D consecutive *PSD*s, $L = D$, the number of compare operations lessens to about k operation per λ step. In any case, if the current noisy speech power spectrum

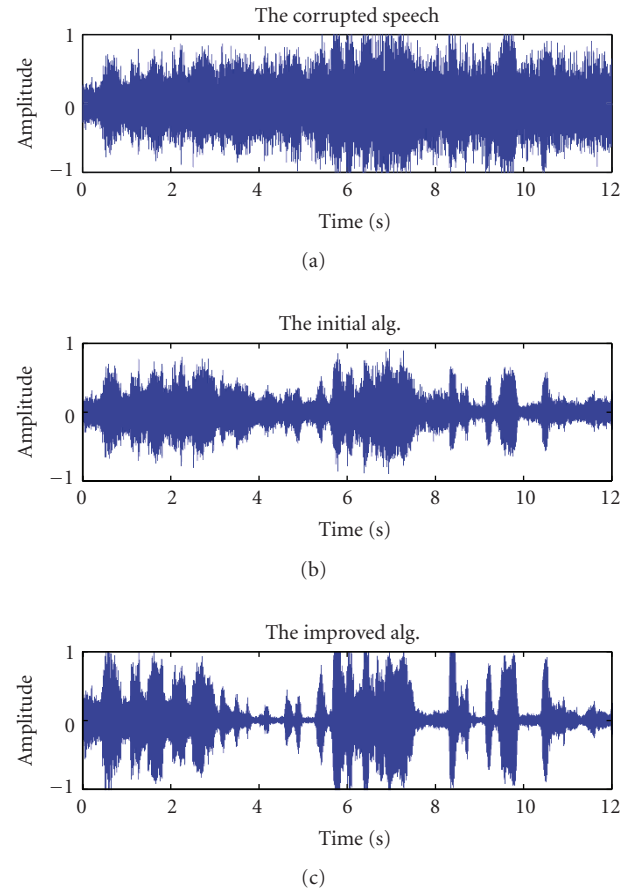

FIGURE 5: Speech signal corrupted with angle grinder noise (a), the initial method produced signal (b), and our modified de-noising method output (c).

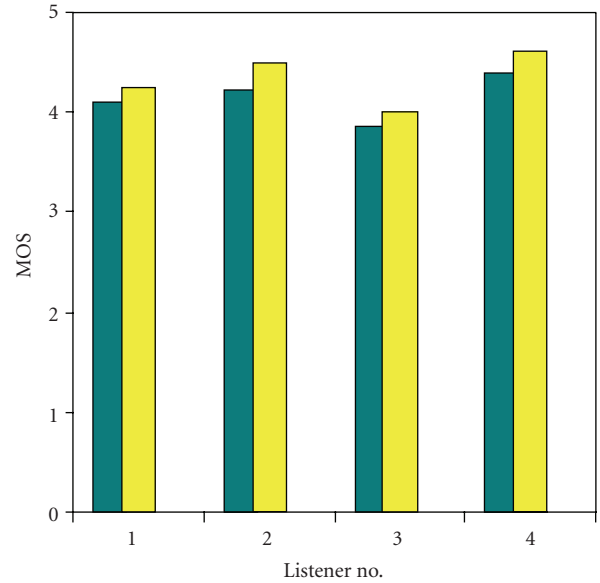

FIGURE 6: Comparison of the perceptual quality of the enhanced speech signals (vertical) by 4 listeners (Horizontal), the dark column: the initial method, and the light column: the modified method.

Table 1: The five-point scale in the Mean Opinion Score.

Rating	Speech quality	Levels of distortion
5	Excellent	Imperceptible
4	Good	Just perceptible but not annoying
3	Fair	Perceptible and annoying
2	Poor	Annoying but not objectionable
1	Unsatisfactory	Objectionable

is smaller than $P_{D\min}(\lambda_D, k)$, the noise power is updated immediately:

$$P_{D\min}(\lambda_D, k) = \min\{P_x(\lambda, k), P_{D\min}(\lambda_D, k)\}, \quad \lambda > \lambda_D. \tag{12}$$

However, in case of increase in noise power in the current frame, the update of the noise estimate is delayed by more than D spectral frames.

The estimate of $P_{D\min}(\lambda_D, k)$ suffers from bias toward lower values that has to be compensated

$$P_{Dn}(\lambda_D, k) = \delta_{\min} P_{D\min}(\lambda_D, k). \tag{13}$$

In case of a relatively white $x(n)$, bias compensation equations have been derived in [18, 33], with the one in [33] being as follows:

$$\delta_{\min}(\lambda_D, k) \approx 1 + (D-1)\frac{\hat{\text{var}}\{P_x(\lambda, k)\}}{\hat{\sigma}_n^4(\lambda_D - L, k)}, \tag{14}$$

where $\lambda_D - L$ indicates the time of the previous P_{Dmin} estimation. The equation indicates that the compensation constant is a function of time, λ and frequency bin, k. However, its exact value will not be optimal for nonstationary situations. Deriving an average value, occasionally, and using it are a remedy that circumvents its computational costs and fits its nonoptimal value.

Incorporating the temporal specs of angle grinder noise in the algorithm has been elaborated in Section 5.2 while employing the frequency specs of noise power has been addressed in Section 5.3.

5.3. Fast Adapting Noise Estimation. To compensate the noise estimation delay, when the noise power jumps, the division of a *D-PSD* block into *C*-weighted *M-PSD* block is considered ($D = C \times M$). It reduces the computational complexity and makes the adaptation faster [18]. The decomposition of the *D-PSD* block into C subblocks has the advantage that a new minimum estimate is available after already M samples without a substantial increase in operations.

The computation steps start with the calculation of the spectral minimum of the first *M frame* spectral minimum as follows:

$$P_{M\min}(\lambda_D + M, k) = \min\{P_x(\lambda_D + M - j, k)\}, \quad j = 0 \cdots M - 1. \tag{15}$$

Then, $P_{M\min}$ for each of the other next M frames is determined. After the calculation of a set of C number of $P_{M\min}$, the next *D-PSD* spectral minimum is derived as follows:

$$P_{D\min}(\lambda_D, k) = \min\{P_{M\min}(\lambda_D - i \times M, k)\}, \quad i = 0 \cdots C - 1. \tag{16}$$

D must be large enough to bridge any peak of speech activity, but short enough to follow nonstationary noise variations. Experiments with different speakers and modulated noise signals have shown that window lengths of approximately 0.8 s–1.4 s give good results [18].

Now, in case of increasing noise power in the current frame, the update of the noise estimate is delayed by $D + M$ spectral frames. To speed up the tracking of the noise spectral minimum, an increase in the importance of the current subframe, with respect to the other past subframes is proposed

$$P_{D\min}(\lambda_D, k) = \min\{\delta_i P_{M\min}(\lambda_D - i \times M, k)\}, \quad i = 0 \cdots C - 1, \tag{17}$$

where δ is a look-ahead constant with $\delta_i \leq \delta_{i-1}$. At the simplest case we have $\delta_i = 1$. Also, for having an accurate noise spectral minimum estimation when a jump occurs in noise power, we modify (12) as follows:

$$P_{D\min}(\lambda_D, k) = \min\{P_x(\lambda, k), \xi P_{M\min}(\lambda_D + i \times M, k)\}, \tag{18}$$

where ξ is the relation-ahead parameter that is related to the segmental NSNR and $\lambda_D + i \times M < \lambda < \lambda_D + (i+1) \times M$. At the simplest situation we set $\xi = 1$. With increasing the value of δ and ξ, the algorithm can track *nonstationary* noises well and the upper bound limit is preventing speech distortions. The above provisions are in close tie with the temporal specs of noise spectrum. In case of angle grinder, change in working conditions from nonengaged (stationary noise) to start of engagement (jump in noise power) to engaged (*nonstationary*) with part and vice versa shapes the dependency of the spectrum to time.

5.4. Multiband Fast Adapting Noise Spectral Estimation. In the case of angle grinder noise, the segmental SNR of high frequency band is significantly lower than the SNR of low frequency band; it implies that their noise variance is different. Another important point that should be considered here is that the high-energy first formant of vowels rests approximately on the frequency band between 400 and 1000 Hz. As a result, this band is not so much susceptible to noise spectrum coarse estimation. On the other hand, the upper frequency band that consonants occupy, the noise spectral estimate should be as precise as possible; otherwise, the intelligibility of speech is impaired. For these reasons, to enhance the performance of our algorithm, we divide the overall spectrum into four regions (0–400 Hz, 400–600 Hz, 600–1000 Hz, and above), and in compliance with (14), separate values for δ and ξ are assigned to each of them. This is somehow similar to the study in [36] regarding colored noise. By this technique, diverse sensitivities in tracking

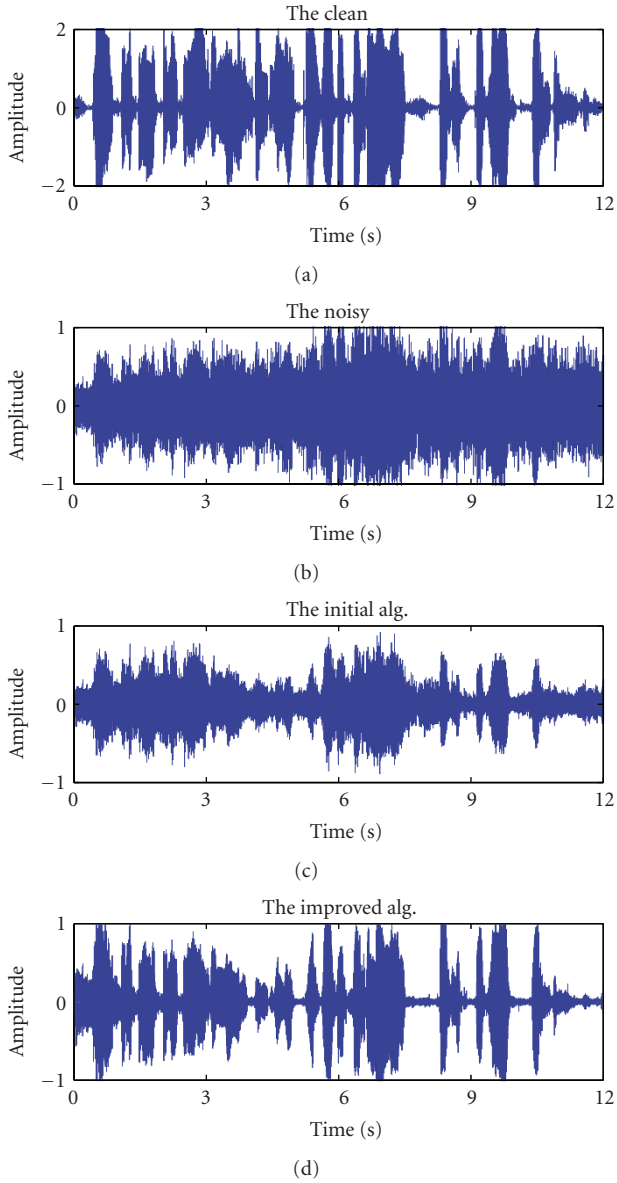

Figure 7: Waveform of the clean, corrupted and enhanced speech signal.

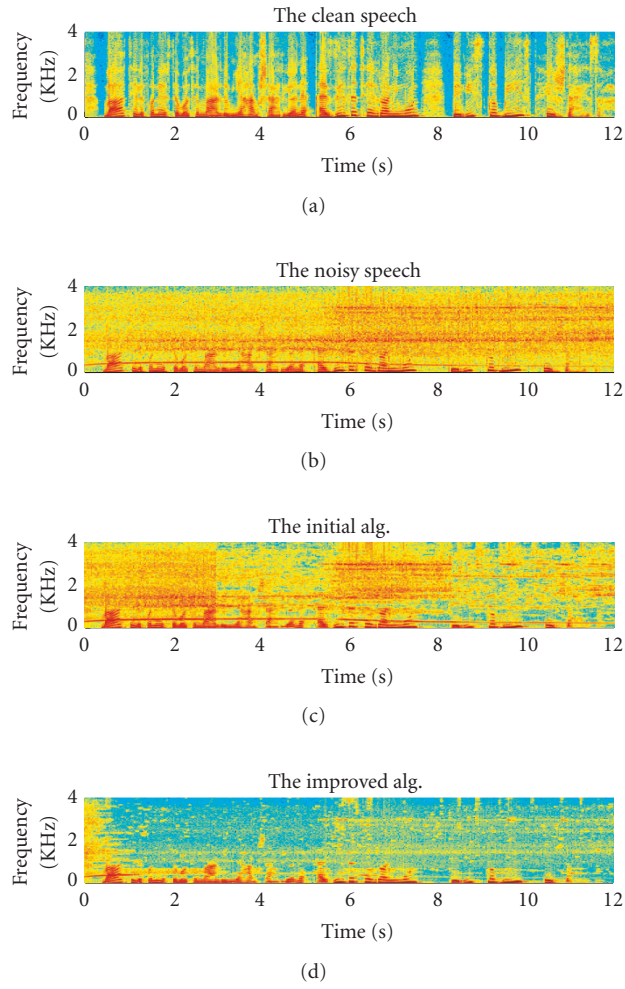

Figure 8: Spectra of the clean, corrupted, and enhanced speech.

nonstationary noise in the different frequency bands are employed. Hence, it is expected that reduction in the speech distortion and increases in the SNR of the processed speech are achieved. For good performance, lower values for δ and ξ in the lower bands are suggested.

6. Performance Evaluation

In order to evaluate the performance of any speech enhancement algorithm, it is necessary to have reliable and appropriate means, based on which the quality and intelligibility of the processed speech can reliably and fairly be quantified. The measures are divided in two groups, objective and subjective measures.

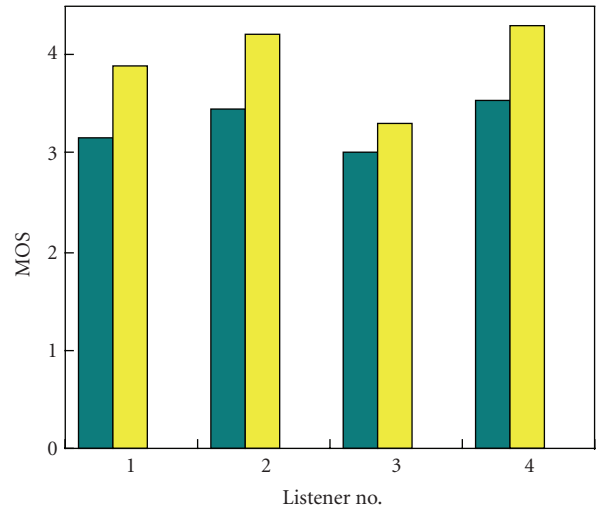

Figure 9: Comparison of the perceptual quality of the enhanced speech signals (vertical) by 4 listeners (horizontal), the dark column: the initial method, and the light column: the modified method.

Table 2: Average of SNR and IS values obtained from 24 male and female speech samples.

		Angle grinder noise (nonengaged)			Angle grinder noise (engaged)		
SNR	in	0	5	10	0	5	10
Seg SNR	in	−1.3	1.83	4.81	−1.15	0.62	3.16
	the initial	3.7	5.70	7.36	1.76	3.21	5.07
	the proposed	4.72	5.82	7.36	2.60	3.6	5.11
SNR_{fw}	in	−11.5	−7.1	−2.95	−13.2	−10.1	−6.01
	the initial	−2.29	0.75	3.26	−6.14	−3.36	0.01
	the proposed	1.8	3.04	4.01	−1.58	0.17	2.12
Seg IS	in	2.05	1.38	0.89	3.64	3.15	2.50
	the initial	1.78	1.31	1.02	2.74	2.57	2.11
	the proposed	0.97	0.69	0.58	2.37	2.03	1.63

6.1. Objective Measures. Segmental SNR is one of the most famous objective measures that is defined by [21]

$$\mathrm{SNR}_M = 10\log\left[\frac{\sum_{k=b_i}^{e_i}|X_M(k)|^2}{\sum_{k=b_i}^{e_i}\left|S_M(k)-\hat{S}_M(k)\right|^2}\right], \tag{19}$$

where $S_M(k)$ and $\hat{S}_M(k)$ are the clean and estimated speech in frame M, respectively.

The other method for calculating SNR is based on a frequency-weighting scheme. This measure better reflects the human auditory system. It is called the Frequency-weighted segment-based SNR ($\mathrm{SNR}_{\mathrm{fw}}$) and is defined by

$$\mathrm{SNR}_{\mathrm{fw}} = \frac{1}{M}\sum_{\lambda=0}^{M-1}\left[\frac{\sum_{k=1}^{N}\alpha_k\times 10\log[(E_s(\lambda,k))/(E_{s-\hat{s}}(\lambda,k))]}{\sum_{k=1}^{N}\alpha_k}\right], \tag{20}$$

where $E_s(j,n)$ and $E_{s-\hat{s}}(j,n)$ denote the short-term signal and noise energy in one of the M frames (index by j), respectively, and the weight α_k is applied to each of the N frequency band indexed by k.

Itakra-Saito (IS) distance is another objective measure that is usually used and has high degree of correlation with the subjective measure ($r = 0.59$) [37]. It performs a comparison between spectral envelopes (all-pole parameters) and that is more influenced by a mismatch in formant location than in spectral valleys. The minimum value of IS corresponds to the best speech quality [27, 29–32, 36, 38]. We use the mean of IS measure that is defined as

$$d(c_1,c_2) = 0.5\left(10\log\frac{c_1R_2c_1'}{c_2R_2c_2'} + 10\log\frac{c_2R_1c_2'}{c_1R_1c_1'}\right), \tag{21}$$

where c_1 and c_2 are the linear prediction coefficient vectors of the clean and enhanced speech segments, respectively. R_1 and R_2 are the Toeplitz autocorrelation matrices of the clean and enhanced speech segment, respectively.

Perceptual Evaluation of Speech Quality (PESQ) enjoys high degree of correlation with the subjective measures ($r = 0.9$) but is one of the most computationally complex of all [39].

6.2. Subjective Measure. In the subjective measure test, the quality of an utterance is evaluated by the opinion of listeners. One of the most often used tests is Mean Opinion Score (MOS), in which listeners rate the speech quality on a five-point scale, according to Table 1.

7. Experimental Setup and Results

Simulations were carried out using 24 Iranian males and females pieces of speeches. Speech samples are recorded in the presence of angle grinder noise in (1) engaged, and (2) non-engaged modes. Signals are sampled at 8 KHz.

7.1. Adaptive Notch Filter. The algorithm worked in canceling pure simulated sine signals, but its performance regarding angle grinder noise was not acceptable. Even though, there are distinct peaks in the spectrum of the angle grinder noise, and the algorithm is able to canceling them; the SNR of the processed signal is not acceptable to be applicable in the headset design. In fact, 1 db improvement in SNR does not satisfy what is really needed.

Further analysis of the noise indicates that the quasiperiodic part of the noise does not carry enough percentage of the noise energy, to the extent that by its removal major improvement occurs. Therefore, other methods of denoising must be considered.

7.2. Fast Adaptive Spectral Subtraction. Signal is framed with an $N = 256$ samples hamming window with *50*% overlap, $R = 128$. In the noise estimation section, the time interval for finding the minimum of noisy speech spectrum is considered 0.72 s, and the number of spectral frames, D, is calculated as follows:

$$\frac{(D-1)R+N}{f_s} = 0.72\,\mathrm{s}, \tag{22}$$

where f_s is the sampling frequency. The $D = 44$ spectral frames is divided into 4 sections each with 11 spectral frames. Then, the estimate of the noise using the modified estimator is computed. We set the values $\delta_1 = 1.01$, $\delta_2 = 1.02$, $\delta_3 = 1.03$,and $\xi = 1.1$ based on the experimental results. Using spectral subtraction with oversubtraction parameter

Table 3: δ and ξ for each of the frequency bands.

	1 Hz $\leq k <$ 400 Hz	400 $\leq k <$ 600 Hz	600 $\leq k <$ 1 KHz	1 KHz $\leq k$
δ_1	1	1	1	1
δ_2	1.01	1.07	1.03	1.1
δ_3	1.05	1.08	1.09	1.12
ξ	1.02	1.1	1.03	1.13

$\alpha_0 = 4$ and spectral floor $\beta = 0.002$, the clean speech in each FFT subwindow is obtained and with taking inverse Fourier transform and overlap and add method, the estimated clean speech signal in the time domain is derived.

Increase in the spectral floor parameter results in residual noise contraction and inversely speech signal distortion. Therefore, an appropriate floor constant (e.g., $\theta = 0.03$) has to be set for the processed signal. As a result, a considerable reduction in the musical noise is gained.

Figure 4 shows one bin, k, of the average smoothed *PSD* of the noisy speech signal, the original noise, the estimated noise by the initial method and the one produced by our improved algorithm. Our method has clearly followed the original noise spectrum. By setting δ and ξ to one, the results tend to the one of the initial method.

Figure 5 shows a piece of speech signal corrupted with a nonstationary angle grinder noise at 0 db SNR, the processed signal by the initial algorithm and by our improved algorithm. It is seen that the proposed algorithm can reduce the noise truly, and the amount of the residual noise is very low.

Table 2 compares the results obtained from averaging SNR and IS distance measures from the processed 24 male and female speech samples. According to Table 2, the value of mean SNR in the proposed algorithm is increased and the mean IS distance is considerably decreased, especially when speech is corrupted with highly nonstationary noise and *SNR* is low. The objective results show superiority of our modified algorithm to the initial algorithm achievements.

To do the subjective test, 3 speech signal samples, each with length 6 Sec, were corrupted with the engaged angle grinder noise under various SNRs. The processed speeches are scored by four listeners. Figure 6 shows the average results gathered from each listener. The dark column is related to the initial method, and the light column is related to our modified method.

As it is shown, the processed speech with the modified algorithm has better perceptual quality than that of the initial algorithm.

7.3. Multi Band Fast Adapting Spectral Subtraction. In this test, the time interval for finding minimum of the noisy speech spectrum is set to 1.5 s:

$$\frac{(D-1)R+N}{f_s} = 1.5\,\text{s} \Longrightarrow D = 92, \tag{23}$$

where $N = 256$ is the time window length. With 50% overlapping, R is 128. The $D = 92$ spectral frame is subdivided into *4* sections of each with 23 spectral frames. Then, the estimate of the noise using the modified estimator is conducted. Based on the experiments, the values of δ and ξ in (17) and (18) in each four bands are set as indicated in Table 3.

As you noticed, different values have been set for each of the 4 frequency bands (low: 1–400 Hz, middle: 400–600 Hz, 600–1000 Hz and above). This accounts for the different noise power in each section of the angle grinder noise spectrum. Using spectral subtraction with oversubtraction parameter $\alpha_0 = 4$ and spectral floor $\beta = 0.002$, the clean speech in each FFT subwindow is obtained. By using Inverse Fourier Transform and Overlap and Add method, the estimated clean speech signal in the time domain is derived. Since with increasing the spectral floor, the residual noise would decrease at the cost of speech signal distortion, we use a time floor constant of $\theta = 0.03$. As a result, a considerable reduction in the musical noise is achieved.

Figures 7 and 8 show the waveform and spectra of a female speech signal corrupted with a nonstationary angle grinder noise at 0 db SNR, and the processed signal by the initial algorithm and the output of the modified multi band algorithm proposed here. It is viewed that the proposed algorithm can reduce the noise truly and the amount of the residual noise is very low. This can be verified better by listening to the pieces of speeches.

Table 4 shows the results obtained from the average of SNR, IS distance and PESQ measures for the improved method in comparison with the initial method. The test was enhancement of 24 male and female speech samples corrupted with noises with various *SNR*s. According to the Table 4, the values of SNR and the PESQ in the proposed algorithm have been increased and the IS distance is considerably decreased, especially for low SNR samples. The objective results show the advantage of our modified algorithm performance versus the initial algorithm results.

To do the subjective test, 24 speech signal samples each with 6-sec-length were corrupted with the engaged angle grinder noise with various SNRs (0 db to 15 db). The processed speeches are scored by four listeners. Figure 9 shows the average results gathered from each listener. The dark column belongs to the initial method, and the light column is related to our improved method. As it is shown, the processed speech with the modified algorithm has better perceptual quality than that of the initial algorithm.

7.4. Overall Assessment. Comparing the contents of Table 2 and Table 4 reveals the outcome gained during this study. In the 0 db *SNR* case, the worst case analyzed here, Table 2 indicates that the method has achieved 2.6 db improvement. The same case in Table 4 shows 6.2 db increase in segmental SNR. Meaning that multiband algorithm is more fit to the case than the single frequency band algorithm. The effectiveness of the algorithm is more noticed in low SNR situations than in moderate SNR cases.

Table 4: The mean SNR, PESQ, and IS values obtained from enhancing 24 noisy male and female speech samples at our experiments for the proposed method compared to the other methods for various SNRs.

		non engaged			engaged		
Input SNR		0	5	10	0	5	10
Seg_SNR	In	−1	2.4	6.2	−1.2	1.56	4.49
	The initial	3.7	6	8.1	1.7	3.94	5.94
	The improved	5.5	6.3	6.9	6.2	7.22	8
SNR_fw	In	−9	−6	−1	−13	−8.5	−4
	The initial	−1	1.3	4.3	−6.2	−2	1.45
	The improved	3	3.8	4.8	2.2	3.8	5.1
PESQ_mos	In	1.4	1.6	1.8	1.52	1.68	1.92
	The initial	1.5	1.9	2.2	1.29	1.62	1.96
	The improved	2	2.3	2.4	1.93	2.19	2.4
IS	In	2.1	1.3	0.7	3.65	2.91	2.22
	The initial	1.7	1.2	0.9	2.77	2.43	1.91
	The improved	0.6	0.5	0.4	1.63	1.42	1.22

8. Conclusion

In this paper, the spectral subtraction method was used to reduce nonstationary angle grinder noise from speech signal. A modified noise estimation algorithm with rapid adaptation for tracking sudden variations in noise power was proposed, and its performance was checked using both objective and subjective measures. It was shown that, the proposed algorithm using multiband weighted subwindow behaves faster and renders more accurate estimate of nonstationary noise and provides a processed signal with minimum musical noise and speech distortion. More works are underway using other appropriate methods. Our challenge is obtaining high quality denoised speech under low SNR situations.

Acknowledgment

This work has been partially supported by the Shahed University research office (SURO), Tehran, Iran.

References

[1] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 27, no. 2, pp. 113–120, 1979.

[2] P. Vary and M. Eurasip, "Noise suppression by spectral magnitude estimation-mechanism and theoretical limits," *Signal Processing*, vol. 8, no. 4, pp. 387–400, 1985.

[3] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proceedings of the IEEE*, vol. 67, no. 12, pp. 1586–1604, 1979.

[4] R. J. McAulay and M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 2, pp. 137–145, 1980.

[5] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 6, pp. 1109–1121, 1984.

[6] G. Iliev and K. Egizarian, "Adaptive system for engine noise cancellation in mobile communications," *Automatica*, vol. 3-4, pp. 137–143, 2004.

[7] Y. Hu and P. C. Loizou, "Speech enhancement based on wavelet thresholding the multitaper spectrum," *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 1, pp. 59–67, 2004.

[8] A. Mouchtaris, J. Van Der Spiegel, P. Mueller, and P. Tsakalides, "A spectral conversion approach to single-channel speech enhancement," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 4, pp. 1180–1193, 2007.

[9] K. K. Paliwal and A. Basu, "A speech enhancement method based on Kalman filtering," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP '87)*, pp. 177–180.

[10] N. Ma, M. Bouchard, and R. A. Goubran, "Perceptual Kalman filtering for speech enhancement in colored noise," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '04)*, vol. 1, pp. 717–720, May 2004.

[11] C. H. You, S. Rahardja, and S. N. Koh, "Perceptual Kalman filtering speech enhancement," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '06)*, vol. 1, pp. 461–464, May 2006.

[12] F. Mustière, M. Bouchard, and M. Bolić, "Low-cost modifications of Rao-Blackwellized particle filters for improved speech denoising," *Signal Processing*, vol. 88, no. 11, pp. 2678–2692, 2008.

[13] G. Yu, S. Mallat, and E. Bacry, "Audio denoising by time-frequency block thresholding," *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 1830–1839, 2008.

[14] Y. Ephraim, "Statistical-model-based speech enhancement systems," *Proceedings of the IEEE*, vol. 80, no. 10, pp. 1526–1554, 1992.

[15] D. Y. Zhao and W. B. Kleijn, "HMM-based gain modeling for enhancement of speech in noise," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 3, pp. 882–892, 2007.

[16] R. C. Hendriks, R. Heusdens, and J. Jensen, "An MMSE estimator for speech enhancement under a combined stochastic-deterministic speech model," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 2, pp. 406–415, 2007.

[17] Q. Yan, S. Vaseghi, E. Zavarehei et al., "Formant tracking linear prediction model using HMMs and Kalman filters for noisy speech processing," *Computer Speech and Language*, vol. 21, no. 3, pp. 543–561, 2007.

[18] R. Martin, "Spectral subtraction based on minimum statistics," in *Proceedings of the 17th European Signal Processing Conference*, pp. 1182–1185, 1994.

[19] G. Doblinger, "Computationally efficient speech enhancement by spectral minima tracking in subbands," in *Proceedings of the 4th European Conference on Speech Communication and Technology (EUROSPEECH '95)*, pp. 1513–1516, Madrid, Spain, September 1995.

[20] H. G. Hirsch and C. Ehrlicher, "Noise estimation techniques for robust speech recognition," in *Proceedings of the 20th International Conference on Acoustics, Speech, and Signal Processing (ICASSP '04)*, pp. 153–156, Detroit, Mich, USA, May 1995.

[21] L. Lin, W. H. Holmes, and E. Ambikairajah, "Subband noise estimation for speech enhancement using a perceptual Wiener

filter," in *Proceedings of the IEEE International Conference on Accoustics, Speech, and Signal Processing (ICASSP '03)*, pp. 80–83, Hong Kong, April 2003.

[22] I. Cohen and B. Berdugo, "Noise estimation by minima controlled recursive averaging for robust speech enhancement," *IEEE Signal Processing Letters*, vol. 9, no. 1, pp. 12–15, 2002.

[23] S. Rangachari, P. C. Loizou, and Y. Hu, "A noise estimation algorithm with rapid adaptation for highly non-stationary environments," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '04)*, pp. 305–308, May 2004.

[24] N. Fan, J. Rosca, and R. Balan, "Speech noise estimation using enhanced minima controlled recursive averaging," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '07)*, vol. 4, pp. 581–584, April 2007.

[25] D. Farrokhi, R. Togneri, and A. Zaknich, "Single channel speech enhancement using a 9 Dimensional noise estimation algorithm and controlled forward march averaging," in *Proceedings of the 9th International Conference on Signal Processing (ICSP '08)*, pp. 17–21, October 2008.

[26] V. Stahl, A. Fischer, and R. Bippus, "Quantile based noise estimation for spectral subtraction and Wiener filtering," in *Proceedings of the IEEE Interntional Conference on Acoustics, Speech, and Signal Processing (ICASSP '00)*, vol. 3, pp. 1875–1878, June 2000.

[27] H. M. Goodarzi and S. Seyedtabaee, "Speech enhancement using spectral subtraction based on a modified noise minimum statistics estimation," in *Proceedings of the 5th International Joint Conference on INC, IMS and IDC*, pp. 1339–1343, 2009.

[28] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proceedings of the IEEE*, vol. 67, no. 12, pp. 1586–1604, 1979.

[29] N. W. D. Evans, J. S. D. Mason, W. M. Liu, and B. Fauve, "An assessment on the fundamental limitations of spectral subtraction," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '06)*, vol. 1, pp. 145–148, May 2006.

[30] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '79)*, pp. 208–211, April 1979.

[31] C. Cole, M. Karam, and H. Aglan, "Spectral subtraction of noise in speech processing applications," in *Proceedings of the 40th Southeastern Symposium on System Theory (SSST '08)*, pp. 50–53, New Orleans, LA, USA, March 2008.

[32] P. Krishnamoorthy and S. R. Prasanna, "Modified spectral subtraction method for enhancement of noisy speech," in *Proceedings of the IEEE 3rd International Conference on Intelligent Sensing and Information Processing*, pp. 146–150, Bangalore, India, December 2005.

[33] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 5, pp. 504–512, 2001.

[34] D. R. Brillinger, *Time Series: Data Analysis and Theory*, Holden-Day, New York, NY, USA, 1981.

[35] N. Derakhshan, A. Akbari, and A. Ayatollahi, "Noise power spectrum estimation using constrained variance spectral smoothing and minima tracking," *Speech Communication*, vol. 51, no. 11, pp. 1098–1113, 2009.

[36] S. Kamath and P. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '02)*, vol. 4, pp. 4160–4164, Orlando, US, 2002.

[37] S. V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction*, John Wiley & Sons, New York, NY, USA, 2000.

[38] J. R. Deller Jr., J. H. L. Hansen, and J. G. Proakis, *Discrete-Time Processing of Speech Signals*, IEEE Press, Piscataway, NJ, USA, 2000.

[39] Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 1, pp. 229–238, 2008.