Circuits Syst Signal Process (2018) 37:1243–1267
https://doi.org/10.1007/s00034-017-0605-7

Speech Noise Reduction Algorithm in Digital Hearing Aids Based on an Improved Sub-band SNR Estimation

**Tao Jiang[1,2] · Ruiyu Liang[1,2] ·
Qinqyun Wang[1] · Yongqiang Bao[1]**

Received: 24 October 2016 / Revised: 30 June 2017 / Accepted: 30 June 2017 /
Published online: 11 July 2017

Abstract To improve the speech intelligibility in noisy environments for persons with hearing impairments, a new method for reducing noise, based on improved sub-band signal-to-noise ratio (SNR) estimation, is proposed. First, the input signal is decomposed into several sub-band signals with an analysis filter bank. Then, under the assumption of a Gaussian model, maximum a posterior probability is applied to estimate the information embedded in adjacent frames in each sub-band, which is in the form of a joint probability density function, and the minimum of the noise spectrum is tracked to estimate the noise. Subsequently, the gain of each sub-band, which changes with the noise in the corresponding sub-band, is calculated with a linear proportional gain function. The obtained gains of the sub-bands are multiplied by the sub-band noisy signals to obtain the enhanced sub-band speech signals. Finally, all the sub-band signals are spliced to obtain the estimated speech signals. In this algorithm, the gains are calculated in the time domain, which avoids the process of the inverse Fourier transform and leads to a decrease in computational complexity. Compared with the traditional spectral subtraction and basic Wiener filtering method, the delay in this algorithm is reduced by 40.4 and 60.6%, respectively. It is also compared with

✉ Ruiyu Liang
lly1711@163.com

Tao Jiang
18351966578@163.com

Qinqyun Wang
wangqingyun@vip.163.com

Yongqiang Bao
jybyq@163.com

[1] School of Communication Engineering, Nanjing Institute of Technology, Nanjing 211167, China

[2] School of Information Science and Engineering, Southeast University, Nanjing 210096, China

Birkhäuser

the modulation depth integrated into hearing aids under an experimental simulation and a real scenario. The results indicate that the output SNR is improved by 1 dB under the software simulation and 3.1 dB in the real scenario when the input SNR is set as 10 dB. Compared with the simulation environment, the proposed algorithm only fell by 1.5% in the real scenario. Furthermore, the distance of the logarithmic spectrum and quality of speech perception are improved by 20.6 and 9.3%, respectively.

Keywords Multi-channel hearing aids · Sub-band noise reduction · Sub-band SNR estimation · Time complexity

1 Introduction

Hearing loss is a common chronic disease affecting human life. Long-term hearing impairment not only influences daily communication but also leads to psychological problems and brings a heavy burden to family and society [1,5,13,15,22,24]. The wearing of a digital hearing aid is an effective way of improving hearing for the patients with light–severe hearing loss [5,13]. However, the improvement by such hearing aids will fall sharply if they operate in a complex noise environment. Research shows that hearing aids need to improve the SNR by 10–25 dB to achieve the same intelligibility as ordinary persons [9,21,32].

Currently, there are two classical methods for improving the SNR in a real scenario, directional microphones and noise reduction algorithms [6]. The former, based on the differences between speech and noise, enhances the speech signals in a specific direction with directional microphones or beam-forming technology [19,34,35]. However, this method is not applied to deep ear hearing aids because of the restrictions on the number and size of microphones. The latter separates the speech from the noisy signals using the differences between speech and noise in the time and frequency domains. However, the speech and noise may overlap in these two domains. To solve this problem, many scholars have done intensive research and have proposed some effective methods.

In the algorithms for reducing noise, spectrum subtraction [3] is one of the earliest algorithms. The principle of this method, based on the stability of noise, is to estimate and update the noise spectrum that will be subtracted from the noisy speech spectrum. Finally, the enhanced signals are obtained by calculating the inverse discrete Fourier transform of the signal spectrum, and its phase is still that of the noisy signal. The time delay of this process is concentrated in the Fourier transform and inverse transform. Recently, some scholars have suggested spectral subtraction to enhance the real and imaginary parts of noisy speech at the same time. For the method based on the phase information [36], its PESQ achieved the improvement of 0.04 and 0.07 for input SNR 10 and 5 dB. However, this slight improvement was achieved by sacrificing double computation complexity. For hearing aids, the adopted algorithms request high real-time performance.

Spectral subtraction could be replaced with Wiener filtering [11] based on a statistical model and its improved algorithm [10], which is based on the minimum mean square of the output signal and pure signal to improve the SNR. The optimal filter (transfer function) is calculated in the frequency domain, and the enhanced signal is obtained through the Fourier inverse transform. Therefore, the delay of this method is similar to that of the traditional algorithms. Furthermore, deep learning [17,30] and the wavelet transform [2,16,26,27,29] also demonstrates good performance in noise reduction. However, these algorithms mainly concentrate on the balance between noise suppression and speech distortion, while the hearing aids with low power consumption should take the speech quality and computational complexity into consideration. In the noise reduction algorithms with low computational complexity, one of the earliest noise reduction algorithms uses a linear proportion function [8] instead of a noise reduction function and a loudness compensation function, which is realized on the dB domain and avoids calculating the FFT and IFFT. According to the above-mentioned analysis, noise reduction with multi-channels, based on modulation depth [12], was proposed and has been widely used in hearing aids. This model contains two parts: tracking the energies of speech and noise and calculating the noise attenuation function, whereas the SNR of the sub-band is calculated through the envelope of the noise and speech signal that is simply smoothed only once. Thus, there is a great deviation in the SNR estimation, which will affect the background noise suppression.

To solve the above problems, a new algorithm for estimating the sub-band SNR in the frequency domain is proposed, which is combined with the linear proportional gain function in the time domain to design the sub-band noise model. First, a decomposition filter bank is applied to separate the noisy signal into several sub-band signals. Then, on the assumption that speech and noise signals are subject to a Gaussian distribution with zero mean, the pure speech signal hidden in the noisy signal is estimated with a maximum posterior probability that contains the relevant information between frames in the form of a joint probability density function. Subsequently, the SNRs of the sub-bands are estimated with the covariance of adjacent frames in the corresponding sub-band. Finally, each sub-band signal is spliced to obtain the enhanced speech signal, which is suppressed by a linear proportional gain function. As shown in the real auditory scene and software simulation experiments, the performance of the proposed algorithm is relative optimum in consideration of the computational complexity and noise reduction performance. This system has some inherent advantages, such as most notably reduced computational complexity and the good performance for some lowest power dedicated processor. So, it can be real-time-implemented and achieve automatic noise reduction for different SNRs. Compared with the noise reduction based on modulation depth [12], the SNR is improved by 1 dB in the software simulation and 3.1 dB in the real scenario. When the input SNR is set as 10 dB, and noise type is set white noise, the LSD and PESQ are improved by 20.6 and 9.3%, respectively. Although the computing time is slightly increased, it is still reduced by 40.4% compared to sub-band spectral subtraction [18] and by 60.6% compared to the Winner filtering bank [10].

2 Algorithm

2.1 Over-Sampling Filter Banks

In the multi-channel noise reduction algorithm, noisy speech is divided into a number of frequency bands according to certain rules, and then, the noise reduction processing is performed in each frequency band. Bands can be divided evenly [33] or not evenly [4]. Because the human ear perception of different frequency bands is nonlinear, nonlinear division is more common. Figure 1 shows the structure of nonlinear over-sampling filter banks. And Table 1 shows the results of band division.

Here, M in the figure is the number of sub-bands, generally taken to be 12, 16 or 20, which is used in the Danish Adcon Safari series of hearing aids. As we know, the more sub-bands the algorithm has, the more complexity and power the system consumes. In most digital hearing aids platforms, because 16 sub-bands is a compromising choice between the performance and the complexity, the number of sub-bands selected in this paper is 16, which means that noisy speech is decomposed into 16 sub-band signals. For the same performance, FIR would consume more time than IIR in embedded systems. So, the filter banks use sixth-order IIR filters, with a pass-band ripple of 0.5 dB.

From the figure, the original speech signal is firstly passed through an anti-aliasing filter bank $h_m(n)$, and the decomposed sub-band signals $u_m(n)$ $(m = 0, 1, \ldots, M-1)$ are obtained. c_m is the lifting sampling coefficient of the mth sub-band, and $c_m \leq M$; because frequency band division was not evenly, c_m should meet Eq. (1) according to band-pass sampling theory to avoid frequency aliasing

$$c_m < \frac{1}{b_m} f_h, \quad m = 0, \ldots, M-1 \tag{1}$$

where b_m represents the width of mth sub-band and f_h represents the maximum frequency. As shown in Table 1 and Eq. (1), specific values are shown in Table 2. $x_m(n)$ is the signal after down-sampling. $g_m(n)$ represents the gain of each sub-band, which is defined according to the SNR and the compensation function; $y_m(n)$ is the over-sampling signal of $v_m(n)$. The final output signal $y(n)$ can be obtained by adding

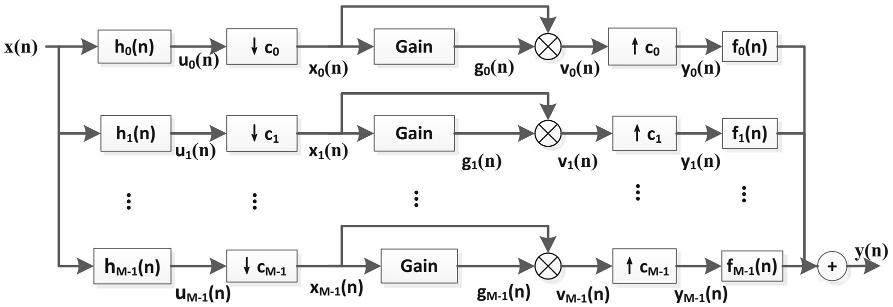


Fig. 1 Structure of the over-sampling filter

Table 1 Frequency band division

Channels (Hz)	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
f_1	1	90	187	312	437	625	875	1125	1375	1625	1875	2250	2750	3250	3750	5500
f_2	90	187	312	437	625	875	1125	1375	1625	1875	2250	2750	3250	3750	5500	7999

Table 2 Down- and up-sampling coefficients

Channel/m	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$c_m\ (n)$	10	10	8	8	6	6	6	6	6	6	4	3	3	3	2	2

those signals that are generated when $y_m(n)$ is passed through the comprehensive filtering.

2.2 Sub-band SNR Estimation Algorithm

For the noise reduction algorithm based on the adaptive gain, a reasonable estimation of the SNR is very important for the performance of noise reduction. Therefore, an improved SNR estimation algorithm for sub-bands is proposed. First, based on the maximum a posteriori probability theory, an algorithm derives the representation of pure speech, so the problem is transformed into the power estimation problem of each frequency point signal. Then, the covariance matrix of adjacent speech frames is constructed, and the power of the speech signal is obtained. Finally, a noise spectrum tracking method is used to estimate the noise power, and the sub-band SNR estimation is realized.

The signal of each sub-band can be expressed as:

$$x_m(n) = s_m(n) + d_m(n)(m = 0, \ldots, M - 1) \tag{2}$$

Within this signal, $s_m(n)$ is the desired pure speech, $d_m(n)$ is the noise signal, and $x_m(n)$ represents the original noisy signal. Each sub-band signal is framed, and the FFT of each frame signal is calculated.

$$\mathbf{X}_m(k, i) = \mathbf{S}_m(k, i) + \mathbf{D}_m(k, i)(i = 0, \ldots, J - 1) \tag{3}$$

In the formula, k represents frequency points, i represents the frame number, J represents the total number of frames, and $\mathbf{X}_m(k, i)$, $\mathbf{S}_m(k, i)$ and $\mathbf{D}_m(k, i)$ represent the FFT of $x_m(n)$, $s_m(n)$ and $d_m(n)$, respectively. They can be expressed as exponential forms: $\mathbf{X}_m(k, i) = X_m(k, i)\mathrm{e}^{j\theta_X}$, $\mathbf{S}_m(k, i) = S_m(k, i)\mathrm{e}^{j\theta_S}$ and $\mathbf{D}_m(k, i) = D_m(k, i)\mathrm{e}^{j\theta_D}$, respectively. Because the ear is not sensitive to the phase, the amplitude estimation of the speech signal is the key to the algorithm.

Assuming that speech signal $S_m(k, i)$ and noise signal $D_m(k, i)$ both obey Gaussian distributions with zero mean, then

$$S_m(k, i) = \frac{1}{\sqrt{2\pi}\sigma_{S_m(k,i)}} \exp\left[\frac{-S_m^2(k, i)}{2\sigma^2_{S_{m(k,i)}}}\right] \tag{4}$$

$$D_m(k, i) = \frac{1}{\sqrt{2\pi}\sigma_{D_m(k,i)}} \exp\left[\frac{-D_m^2(k, i)}{2\sigma^2_{D_{m(k,i)}}}\right] \tag{5}$$

In the formula, $\sigma^2_{S_m(k,i)}$ and $\sigma^2_{D_m(k,i)}$ represent the variance of speech signal $S_m(k,i)$ and noisy signal $D_m(k,i)$. The enhanced speech signal is represented by $\hat{S}_m(k,i)$; then, the error estimation function is defined as [11]

$$d(\varepsilon) = \varepsilon^2 = \left[S_m(k,i) - \hat{S}_m(k,i)\right]^2 \tag{6}$$

Obviously, the model is the minimum mean square error model of the Wiener filtering method [11]. However, the method involves a large number of integral and exponential operations and is not suitable for low-power devices such as hearing aids. Therefore, the improved algorithm is proposed to reduce the computational complexity [20]. Here, $d(\varepsilon)$ is defined as

$$d(\varepsilon) = \begin{cases} 0 & |\varepsilon| < \delta \\ 1 & |\varepsilon| > \delta \end{cases} \tag{7}$$

The risk function of the error function $\Re$ is defined as

$$\Re = E[d(\varepsilon)] = \int \left\{ \int d(\varepsilon) P[S_m(k,i)|W_m(k,i)] dX_m(k,i) \right\} P[W_m(k,i)] dW_m(k,i) \tag{8}$$

where $W_m(k,i) = [X_m(k,i), S_m(k,i-1)]$.

By substituting Eq. (7) into Eq. (8) and minimizing $\Re$, $\hat{S}_m(k,i)$ can then be deduced as [28]:

$$\hat{S}_m(k,i) = \arg \min_{S_m(k,i)} \Re = \arg \max_{S_m(k,i)} P[S_m(k,i)|X_m(k,i), S_m(k,i-1)] \tag{9}$$

According to Bias theory, $P[S_m(k,i)|X_m(k,i), S_m(k,i-1)]$ can be rewritten as:

$$\begin{aligned} &P[S_m(k,i)|X_m(k,i), S_m(k,i-1)] \\ &\quad = \frac{P[X_m(k,i), S_m(k,i-1)|S_m(k,i)] P[S_m(k,i)]}{P[X_m(k,i), S_m(k,i-1)]} \\ &\quad = \frac{P[X_m(k,i)|S_m(k,i)] P[S_m(k,i), S_m(k,i-1)]}{P[X_m(k,i), S_m(k,i-1)]} \end{aligned} \tag{10}$$

In Eq. (10), the denominator $P[X_m(k,i), S_m(k,i-1)]$ and $S_m(k,i)$ are independent. Thus, $\hat{S}_m(k,i)$ can be rewritten as:

$$\hat{S}_m(k,i) = \arg \max_{S_m(k,i)} \{P[X_m(k,i)|S_m(k,i)] P[S_m(k,i), S_m(k,i-1)]\} \tag{11}$$

Assuming the speech signal $S_m(k,i)$ and noise $D_m(k,i)$ are independent, and noise is stationary, then

$$P[X_m(k,i)|S_m(k,i)] = \frac{2}{\sqrt{2\pi}\sigma_{D_m(k,i)}} \exp\left\{ \frac{-[X_m(k,i) - S_m(k,i)]^2}{2\sigma^2_{D_m(k,i)}} \right\} \tag{12}$$

From Eq. (12), the estimation of $P[S_m(k,i), S_m(k,i-1)]$ is the key to obtain an effective estimation of $\hat{S}_m(k,i)$. Firstly, define the adjacent two speech vectors $\mathbf{S}_m^a(k,i)$ as

$$\mathbf{S}_m^a(k,i) = \{S_m(k,i), S_m(k,i-1)\}^{\mathrm{T}} \tag{13}$$

where T represents a transposition. Then, the autocorrelation matrix can be expressed as:

$$\begin{aligned}\mathbf{R}_m(k,i) &= E[\mathbf{S}_m^a(k,i)\mathbf{S}_m^a(k,i)^H] \\ &= \begin{pmatrix} E[S_m^2(k,i)] & E[S_m(k,i)S_m(k,i-1)] \\ E[S_m(k,i)S_m(k,i-1)] & E[S_m^2(k,i-1)] \end{pmatrix} \\ &= \begin{pmatrix} \sigma^2_{S_m(k,i)} & \bar{\lambda}_m(i)\sigma^2_{S_m(k,i-1)} \\ \bar{\lambda}_m(i)\sigma^2_{S_m(k,i-1)} & \sigma^2_{S_m(k,i-1)} \end{pmatrix}\end{aligned} \tag{14}$$

where H represents the conjugate. The derivation of $E[S_m(k,i)S_m(k,i-1)]$ is:

$$\begin{aligned}E[S_m(k,i)S_m(k,i-1)] &= E\left[\frac{S_m(k,i)}{S_m(k,i-1)}S_m^2(k,i-1)\right] \\ &= E[\lambda_m(k,i)S_m^2(k,i-1)] \approx \bar{\lambda}_m(i)\sigma^2_{S_m(k,i-1)}\end{aligned} \tag{15}$$

Thus, $P[S_m(k,i), S_m(k,i-1)]$ can be rewritten as:

$$\begin{aligned}&P[S_m(k,i), S_m(k,i-1)] \\ &\quad= \frac{1}{2\pi\,|\mathbf{R}_m(k,i)|^{1/2}}\exp\left[-\frac{1}{2}\mathbf{S}_m^a(k,i)^{\mathrm{T}}\mathbf{R}_m(k,i)^{-1}\mathbf{S}_m(k,i)\right] \\ &\quad= \frac{1}{2\pi\,S_m(k,i-1)\sqrt{\sigma^2_{S_m(k,i)} - \bar{\lambda}_m^2(i)\sigma^2_{S_m(k,i-1)}}} \\ &\quad\exp\left\{-\frac{S_m^2(k,i)\sigma^2_{S_m(k,i-1)} - 2\bar{\lambda}_m(i)S_m(k,i-1)S_m(k,i)\sigma^2_{S_m(k,i-1)} + S_m^2(k,i-1)\sigma^2_{S_m(k,i)}}{2\left[\sigma^2_{S_m(k,i)}\sigma^2_{S_m(k,i-1)} - \bar{\lambda}_m^2(i)\sigma^2_{S_m(k,i-1)}\right]}\right.\end{aligned} \tag{16}$$

Combining Eqs. (12) and (16), $P[X_m(k,i)|S_m(k,i)]P[S_m(k,i), S_m(k,i-1)]$ can be written as:

$$\begin{aligned}&P[X_m(k,i)|S_m(k,i)]P[S_m(k,i), S_m(k,i-1)] \\ &\quad= \frac{2}{\sqrt{2\pi}\sigma_{D_m(k,i)}}\cdot\frac{1}{2\pi\,S_m(k,i-1)\sqrt{\sigma^2_{S_m(k,i)} - \bar{\lambda}_m^2(i)\sigma^2_{S_m(k,i-1)}}} \\ &\quad\exp\left\{-\frac{S_m^2(k,i)\sigma^2_{S_m(k,i-1)} - 2\bar{\lambda}_m(i)S_m(k,i-1)S_m(k,i)\sigma^2_{S_m(k,i-1)} + S_m^2(k,i-1)\sigma^2_{S_m(k,i)}}{2\left[\sigma^2_{S_m(k,i)}\sigma^2_{S_m(k,i-1)} - \bar{\lambda}_m^2(i)\sigma^2_{S_m(k,i-1)}\right]}\right. \\ &\quad\left. -\frac{[X_m(k,i) - S_m(k,i)]^2}{2\sigma^2_{D_m(k,i)}}\right\}\end{aligned} \tag{17}$$

Defining the exponent part as Z and calculating the partial derivative of $S_m(k,i)$, then

$$\begin{aligned}\frac{\partial Z}{\partial S_m(k,i)} &= \frac{\partial}{\partial S_m(k,i)} \\ &\left\{-\frac{S_m^2(k,i)\sigma^2_{S_m(k,i-1)} - 2\bar{\lambda}_m(i)S_m(k,i-1)S_m(k,i)\sigma^2_{S_m(k,i-1)} + S_m^2(k,i-1)\sigma^2_{S_m(k,i)}}{2\left[\sigma^2_{S_m(k,i)}\sigma^2_{S_m(k,i-1)} - \bar{\lambda}^2_m(i)\sigma^4_{S_m(k,i-1)}\right]} \right. \\ &\quad \left. -\frac{[X_m(k,i)-S_m(k,i)]^2}{2\sigma^2_{D_m(k,i)}}\right\} \\ &= -\frac{S_m(k,i)\sigma^2_{S_m(k,i-1)} - \bar{\lambda}_m(i)S_m(k,i-1)\sigma^2_{S_m(k,i-1)}}{\left[\sigma^2_{S_m(k,i)}\sigma^2_{S_m(k,i-1)} - \bar{\lambda}^2_m(i)\sigma^4_{S_m(k,i-1)}\right]} + \frac{X_m(k,i)-S_m(k,i)}{\sigma^2_{D_m(k,i)}} \end{aligned} \tag{18}$$

Letting $\frac{\partial Z}{\partial S_m(k,i)} = 0$, the estimated enhanced speech signal $\hat{S}_m(k,i)$is expressed as

$$\begin{aligned}\hat{S}_m(k,i) &= \frac{\sigma^2_{S_m(k,i)} - \bar{\lambda}^2_m(i)\sigma^2_{S_m(k,i-1)}}{\sigma^2_{D_m(k,i)} + [1-\bar{\lambda}^2_m(i)]\sigma^2_{S_m(k,i)}} X_m(k,i) \\ &\quad + \frac{\bar{\lambda}_m(i)\sigma^2_{D_m(k,i)}}{\sigma^2_{D_m(k,i)} + [1-\bar{\lambda}^2_m(i)]\sigma^2_{S_m(k,i)}} S_m(k,i-1)\end{aligned} \tag{19}$$

Letting $\gamma_m(k,i) = \frac{\sigma^2_{S_m(k,i)} - \bar{\lambda}^2_m(i)\sigma^2_{S_m(k,i-1)}}{\sigma^2_{D_m(k,i)} + [1-\bar{\lambda}^2_m(i)]\sigma^2_{S_m(k,i)}}$ and $\eta_m(k,i) = \frac{\bar{\lambda}_m(i)\sigma^2_{D_m(k,i)}}{\sigma^2_{D_m(k,i)} + [1-\bar{\lambda}^2_m(i)]\sigma^2_{S_m(k,i)}}$, $\hat{S}_m(k,i)$ can then be written as:

$$\hat{S}_m(k,i) = \eta_m(k,i)S_m(k,i-1) + \gamma_m(k,i)X_m(k,i) \tag{20}$$

If $\gamma_m(k,i)$ and $\eta_m(k,i)$ is set as constant, such as ts and $1-ts$, then the algorithm becomes the smooth method in finding [12], which use the envelope of speech information, but it will still obtain residual background noise. In Eq. (19), if $S_m(k,i)$ and $S_m(k,i-1)$ are independent, then $\bar{\lambda}^2_m(i)$ equals 0, and the model is equivalent to the Wiener filtering model [11]. And if $\sigma^2_{S_m(k,i-1)}$ and $\sigma^2_{S_m(k,i)}$ change slowly ($\sigma^2_{S_m(k,i)} \approx \sigma^2_{S_m(k,i-1)}$), then the pure speech estimation is the same as in the literature [28], which is shown below:

$$\begin{aligned}\hat{S}_m(k,i) &= (1-ts)\cdot S_m(k,i-1) + ts\cdot X_m(k,i) \\ &= \frac{[1-\bar{\lambda}^2_m(i)]\sigma^2_{S_m(k,i)}}{\sigma^2_{D_m(k,i)} + [1-\bar{\lambda}^2_m(i)]\sigma^2_{S_m(k,i)}} X_m(k,i) \\ &\quad + \frac{\bar{\lambda}_m(i)\sigma^2_{D_m(k,i)}}{\sigma^2_{D_m(k,i)} + [1-\bar{\lambda}^2_m(i)]\sigma^2_{S_m(k,i)}} S_m(k,i-1)\end{aligned} \tag{21}$$

In fact, as seen from Eq. (19), once the signal $\hat{S}_m(k,i)$ in the frequency domain was estimated, the time-domain signal $\hat{s}_m(n,i)$ could be obtained by Fourier inversion. However, the Fourier inversion requires a large amount of calculation, which will increase the computational complexity. Moreover, for the power of each frequency point, it is difficult to accurately estimate $\sigma^2_{D_m(k,i)}$, $\sigma^2_{S_m(k,i-1)}$ and $\sigma^2_{S_m(k,i)}$. Additionally, it needs large time-consuming to estimate the variance of each frequency point.

For above reasons, two improved strategies are proposed:

1. With the consistency of the SNR, it is calculated in the frequency domain and signal gain in the time domain. Because the SNR of the signal is consistent, both in the time domain and frequency domain

$$SNR_m(n,i) = 10\log\left[\frac{|\hat{s}_m(n,i)|^2}{|\hat{d}_m(n,i)|^2}\right] = 10\log\left[\frac{|\hat{S}_m(k,i)|^2}{|\hat{D}_m(k,i)|^2}\right]. \tag{22}$$

2. The information of L frequency points are used to estimate the variance of a signal $\sigma^2_{D_m(k,i)}$, $\sigma^2_{S_m(k,i-1)}$ and $\sigma^2_{S_m(k,i)}$. These three variances are used to approximately replace $\sigma^2_{S_m(k,i)}$, $\sigma^2_{S_m(k,i-1)}$ and $\sigma^2_{D_m(k,i)}$. So, $\gamma_m(k,i)$ and $\eta_m(k,i)$ can be written as

$$\gamma_m(k,i) \approx \gamma_m(i) = \frac{\sigma^2_{S_m(i)} - \bar{\lambda}^2_m(i)\sigma^2_{S_m(i-1)}}{\sigma^2_{D_m(i)} + [1-\bar{\lambda}^2_m(i)]\sigma^2_{S_m(i)}} \tag{23}$$

$$\eta_m(k,i) \approx \eta_m(i) = \frac{\bar{\lambda}_m(i)\sigma^2_{D_m(i)}}{\sigma^2_{D_m(i)} + [1-\bar{\lambda}^2_m(i)]\sigma^2_{S_m(i)}} \tag{24}$$

It can now be seen that the estimation of $\sigma^2_{S_m(i)}$, $\sigma^2_{S_m(i-1)}$ and $\bar{\lambda}^2_m(i)$ are the key to the algorithm. After a signal matrix $\mathbf{X}^a_m(k,i) = \{X_m(k,i), X_m(k,i-1)\}^{\mathrm{T}}$ of two adjacent frames is defined, the autocorrelation matrix $\mathbf{R}^x_m(k,i)$ is expressed as:

$$\begin{aligned}\mathbf{R}^x_m(k,i) &= E\left[\mathbf{X}^a_m(k,i)\mathbf{X}^a_m(k,i)^H\right] = \begin{pmatrix} R_1 & R_2 \\ R_3 & R_4 \end{pmatrix} \\ &= \begin{pmatrix} E[X^2_m(k,i)] & E[X_m(k,i)X_m(k,i-1)] \\ E[X_m(k,i)X_m(k,i-1)] & E[X^2_m(k,i-1)] \end{pmatrix} \\ &= \begin{pmatrix} \sigma^2_{S_m(k,i)} + \sigma^2_{D_m(k,i)} & \bar{\lambda}_m(i)\sigma^2_{S_m(k,i-1)} \\ \bar{\lambda}_m(i)\sigma^2_{S_m(k,i-1)} & \sigma^2_{S_m(k,i-1)} + \sigma^2_{D_m(k,i-1)} \end{pmatrix}\end{aligned} \tag{25}$$

The $\mathbf{R}^x_m(i)$ of L frequency points are used to replace the $\mathbf{R}^x_m(k,i)$ of the single frequency point, and Eq. (25) becomes:

$$\begin{aligned}\mathbf{R}^x_m(i) &= \begin{pmatrix} Q_1 & Q_2 \\ Q_3 & Q_4 \end{pmatrix} = \begin{pmatrix} E[X^2_m(i)] & E[X_m(i)X_m(i-1)] \\ E[X_m(i)X_m(i-1)] & E[X^2_m(i-1)] \end{pmatrix} \\ &= \begin{pmatrix} \sigma^2_{S_m(i)} + \sigma^2_{D_m(i)} & \bar{\lambda}_m(i)\sigma^2_{S_m(i-1)} \\ \bar{\lambda}_m(i)\sigma^2_{S_m(i-1)} & \sigma^2_{S_m(k,i-1)} + \sigma^2_{D_m(i-1)} \end{pmatrix}\end{aligned} \tag{26}$$

Q_1, Q_2, Q_3 and Q_4 can be calculated by the following formula:

$$\begin{cases} Q_1 = E[X_m^2(i)] = \frac{1}{L}\sum\limits_{k=0}^{L-1} X_m^2(k,i) \\ Q_4 = E[X_m^2(i-1)] = \frac{1}{L}\sum\limits_{k=0}^{L-1} X_m^2(k,i-1) \\ Q_2 = Q_3 = E[X_m(i)X_m(i-1)] = \frac{1}{L}\sum\limits_{k=0}^{L-1} X_m(k,i)X_m(k,i-1) \end{cases} \tag{27}$$

Equation (26) is joined with Eq. (27), so

$$\begin{cases} \sigma^2_{S_m(i)} = Q_1 - \sigma^2_{D_m(i)} \\ \sigma^2_{S_m(i-1)} = Q_4 - \sigma^2_{D_m(i-1)} \\ \bar{\lambda}_m(i) = Q_2 \Big/ \left[Q_4 - \sigma^2_{D_m(i-1)}\right] = Q_3 \Big/ \left[Q_4 - \sigma^2_{D_m(i-1)}\right] \end{cases} \tag{28}$$

It is clear that the solution of the problem becomes the estimation of the variance $\sigma^2_{D_m(i)}$ and $\sigma^2_{D_m(i-1)}$ of the noise. In the literature, the power spectrum estimation of the noise usually uses the minimum value tracking method [7,14,25,31]. To reduce the computational complexity and storage resources, the search window length is set to a finite length. The specific steps are as follows:

Step 1 Calculate the signal power and smooth it.

$$\sigma^2_{\bar{\bar{X}}_m(i)} = \alpha\sigma^2_{\bar{\bar{X}}_m(i-1)} + (1-\alpha)\sigma^2_{\bar{\bar{X}}_m(i)} \tag{29}$$

Here, α is the smoothing factor and $\alpha = 0.7$. $\sigma^2_{\bar{\bar{X}}_m(i)}$ represents the noise variance, and its calculation method is the same as that for Q_1.

Step 2 Search for the minimum value in the previous u frames. If the power of the current frame signal $\sigma^2_{\bar{\bar{X}}_m(i)}$ is less than the power $\sigma^2_{D_m^{\min}(i-1)}$ of the previous frame noise, then the power of the current frame signal is the same as the power of this frame noise $\sigma^2_{D_m^{\min}(i)}$. Otherwise, search the minimum value from the previous $u-1$ frames to obtain $\sigma^2_{D_m^{\min}(i)}$. That is,

$$\sigma^2_{D_m^{\min}(i)} = \min\left\{\sigma^2_{\bar{\bar{X}}_m(i-1)}, \sigma^2_{\bar{\bar{X}}_m(i-2)}, \ldots, \sigma^2_{\bar{\bar{X}}_m(i-u)}\right\} \tag{30}$$

Here, the value of u is limited, which indicates that searching is performed in the adjacent frames, rather than looking through a number of previous frames.

Step 3 Calculate the probability of the existence of the voice $p_m^s(i)$

$$p_m^s(i) = \beta p_m^s(i-1) + (1-\beta)I_m^s(i) \tag{31}$$

Here, β is the probability update coefficient, and the value is 0.2. $I_m^s(i)$ is the state function that represents the presence of a voice, which is defined as

$$I_m^s(i) = \begin{cases} 1 & \sigma^2_{\bar{X}_m(i)} / \sigma^2_{D_m^{\min}(i)} > \delta \\ 0 & \sigma^2_{\bar{X}_m(i)} / \sigma^2_{D_m^{\min}(i)} \leq \delta \end{cases} \tag{32}$$

where δ is the threshold value, and it takes to be 2.

Step 4 Smooth to obtain the final noise spectrum

$$\sigma^2_{D_m(i)} = \alpha_m(i)\sigma^2_{D_m(i-1)} + [1 - \alpha_m(i)]\sigma^2_{\bar{X}_m(i)}. \tag{33}$$

In the formula, $\alpha_m(i)$ is the smoothing factor, which is defined as

$$\alpha_m(i) = \alpha + (1 - \alpha)p_m^s(i) \tag{34}$$

Finally, the SNR is estimated as

$$SNR_m(i) = 20\lg\left[\frac{|s_m(i)|}{|d_m(i)|}\right] = 20\lg\left[\frac{|S_m(i)|}{|D_m(i)|}\right] \approx 20\lg\left[\frac{|\hat{S}_m(i)|}{|X(i) - \hat{S}_m(i)|}\right] \tag{35}$$

2.3 The Construction of the Gain Function

In digital hearing aids, the noise reduction usually precedes the loudness compensation in signal processing, which is designed to suppress background noise and avoid amplifying the background noise in the subsequent loudness compensation algorithm. The basic function of loudness compensation is to calculate the signal gain of each frequency band according to the audiometric curve of the patient and the sound pressure level (SPL) of the input signal, to thus compensate for the patient's hearing loss in each frequency band, and to enhance the audibility of speech, thereby improving the speech intelligibility of the patient. The function can be expressed by the gain function $g^l_{\mathrm{dB}}(i, m)$, which calculates the gain in the dB domain, and it is then transformed to the amplitude domain through the following formula.

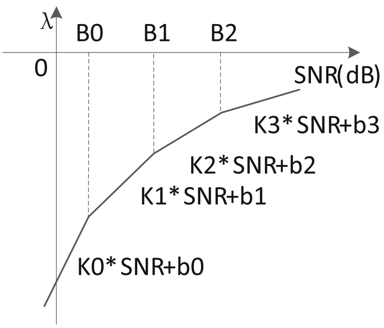
$$g^l(i, m) = 10^{2g^l_{\mathrm{dB}}(i,m)} \tag{36}$$

The modulation depth method of noise reduction commonly used in a hearing aid and it is the same as the loudness compensation in principle, namely calculating the gain function based on the SNR, thus reducing the noise in the signal. The gain function is defined as

$$g_{\mathrm{dB}}(i, m) = \lambda_{im}(SNR)f(\sigma^2_{D_m(i)}) \tag{37}$$

$$g(i, m) = 10^{2\cdot g_{\mathrm{dB}}(i,m)} \tag{38}$$

Fig. 2 Modified linear proportional gain function



In Eqs. (37) and (38), $g_{db}(i,m)$ represents the gain of the ith frame in the mth sub-band in the dB domain, $g(i,m)$ represents the gain in the amplitude domain, and $\lambda_{im}(SNR)$ is an attenuation function with the change of the SNR, the value of which is limited to $[-1, 0]$. When $\lambda_{im}(SNR)$ is 0, the corresponding gain in the amplitude domain is 1, which means that there is no attenuation. If the value of $\lambda_{im}(SNR)$ is close to -1, the signal will obtain the maximum attenuation in the amplitude domain. Early methods were based on the relationship between the linear proportional gain function and the SNR in the sub-band to suppress noise, but some intermittent background noise is still retained, which leads to the user's auditory fatigue. Therefore, an improved strategy is proposed, namely adding a section of attenuation polylines in the parts with low SNR. As shown in Fig. 2, where $K0$, $K1$, $K2$ and $K3$ represent different attenuation slopes in different SNR range. More specifically, in the range of $[0, B0]$, the attenuation curve is relatively steep, which can better restrain residual noise. However, this strategy is not suitable for a weak signal environment. Therefore, the current environmental noise level is added into the algorithm design. For example, when the patient is in a relatively quiet environment, and if the gain function relies solely on the current SNR, the voice signal will be attenuated. Instead, the maximum attenuation function $f(\sigma^2_{D_m(i)})$ is used to control the magnitude of the attenuation. Therefore, the maximum attenuation function is set to the current noise energy $\sigma^2_{D_m(i)}$, so even if the current SNR is low, the speech signal will not decay too much.

2.4 A De-noising Algorithm Based on the Improved Estimation of the Sub-bands' SNR

As the figure shows, the noisy speech signal $\mathbf{x}(n)$ will be divided into several sub-bands through some band-pass filters with various widths, and then, each sub-band will be down-sampled and subjected to the Fourier transformation. The real-time SNR of each sub-band will be calculated based on the estimation of the speech signal and sub-band noise in frequency domain. And according to Eq. (35), $\text{SNR}_m(i)$ can be obtained. Hence, the gain in the time domain can be calculated from the $\text{SNR}_m(i)$ of the sub-bands and then be multiplied with the down-sampled signal. Next, the enhanced version of each sub-band signal by up-sampling the original signal can be obtained, and the enhanced digital signal can be generated by processing the enhanced sub-band

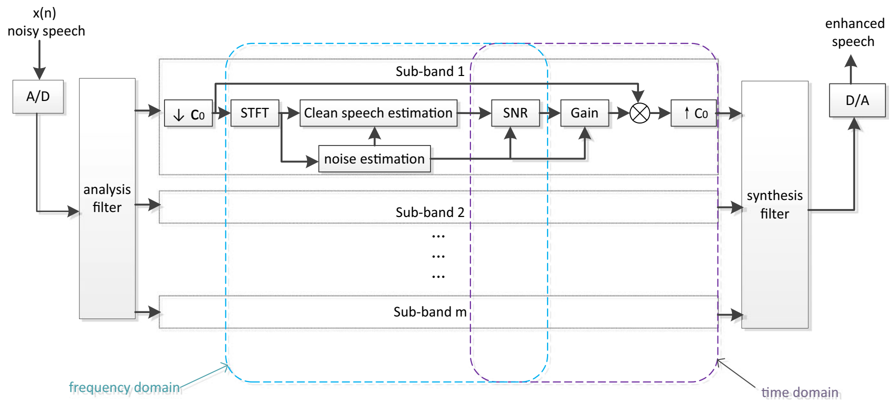


Fig. 3 The flow chart of the de-noising algorithm based on the improved estimation of the sub-bands' SNR

signals with synthetic filters. Finally, the digital signal passes the D/A converter, and the enhanced speech signal is generated. The flow chart of de-noising algorithm based on the improved sub-band SNR estimation is shown in Fig. 3. The specific steps in the calculation are as follows:

1. Divide the input signal into various sub-bands using the band-pass filters with different widths as described in Sect. 2.1.
2. Down-sample each sub-band signal, frame the signal and perform the short-time Fourier transformation.
3. Use the minimum value tracking method to estimate the spectrum of noise in each sub-band with Eq. (33);
4. Estimate the pure speech signal with the algorithm for the pure speech signal of each sub-band through Eq. (19). The parameters are estimated Eq. (28). Due to the previous frame of speech not being available, the estimated frame of speech $\hat{S}_m(k, i-1)$ is used to replace the $S_m(k, i-1)$ in Eq. (19).
5. Calculate the SNR of each sub-band with Eq. (35).
6. For overall consideration of the sub-band SNR and the noise level of the environment, the gain of each sub-band signal with Eq. (38) is calculated.
7. Then, the sub-band signal is multiplied with its respective gain.
8. The final enhanced speech signal can be obtained by up-sampling the result of the last step and then filtering it with the synthetic filters.

3 Simulation Experiments

In order to compare the performance and the calculation of algorithms, some experiments are tested on the traditional spectrum subtraction method [3], sub-band spectrum subtraction method [18], modulation depth method [12], basic Wiener method [11] and improved Wiener method [10]. The performance indexes include the SNR, log spectral distortion (LSD), perceptual evaluation of speech quality (PESQ) and system delay. All the experiments were arranged in a mute room, and the broadcast facility was a loudspeaker array (including 4 loudspeakers and one woofer). The test scene can be simulated by SurroundRouter, a simulation software. The experiment noise is from the NoiseX-92 dataset, which mainly includes white noise, tank noise, babble speech noise and pink noise. The input SNRs were set 0, 5, and 10 dB. Speech files come from the TIMIT speech dataset and some self-recorded speech files with a 16 kHz sample rate. The hearing aid for testing was placed in the center of the loud speaker array. TES-52A is used to calibrate the sound pressure level of the hearing aid. To compare the time delay performance of various algorithms, the length of noisy speech was 0.55, 1.1, and 2.2 s.

Of the targets above, LSD reflects the speech distortion and PESQ can indicate the overall speech quality. There is a high correlation between those two indices and subjective assessment. Usually, the more the LSD decreases, the less the log spectrum distortion is and the less damage done to speech is. Meanwhile, the more the PESQ increases, the better quality of speech. The calculating methods of LSD and PESQ are as follows:

$$LSD = \frac{1}{J}\sum_{l=0}^{J-1}\left\{\frac{1}{N/2+1}\sum_{k=0}^{N/2}\left[10\log_{10}X(k,l) - 10\log_{10}\hat{X}(k,l)\right]^{2}\right\}^{\frac{1}{2}} \tag{39}$$

$$PESQ = 4.5 - 0.1\cdot d_{\mathrm{SYM}} - 0.0309\cdot d_{\mathrm{ASYM}} \tag{40}$$

In Eq. (39), $X(k,l)$ and $\hat{X}(k,l)$ are the short-time Fourier transformation of pure speech and enhanced speech, respectively. N is the frame length, and J is the frame number. The calculations of the d_{SYM} and d_{ASYM} can be found in literature [23].

3.1 Sub-band SNR Estimation Experiments

Under the same experimental conditions as the sub-band speech estimation experiments, this experiment has verified the performance of sub-band SNR estimation, and the results are shown in Fig. 4. When the SNR is above 0 dB, the average bias of the SNR estimation is small, 2.76 dB under white noise and 2.48 dB under pink noise. However, when the SNR is below 0 dB, the estimation will have a large bias, which reaches 5.1 dB under white noise and 4.5 dB under pink noise. The reason for this circumstance is that estimation of the speech signal in silent periods of frames is large. However, it is meaningless for digital hearing aid customers to perform the estimation below 0 dB. That is why signals below 0 dB will be extremely suppressed in the attenuation module.

3.2 Speech Noise Reduction Performance Comparison Experiment

3.2.1 Improvement in the Speech Spectrum

Figure 5 shows the spectrogram comparison of different noise reduction algorithms. The spectrogram describes the relative energy distribution of the speech signal in

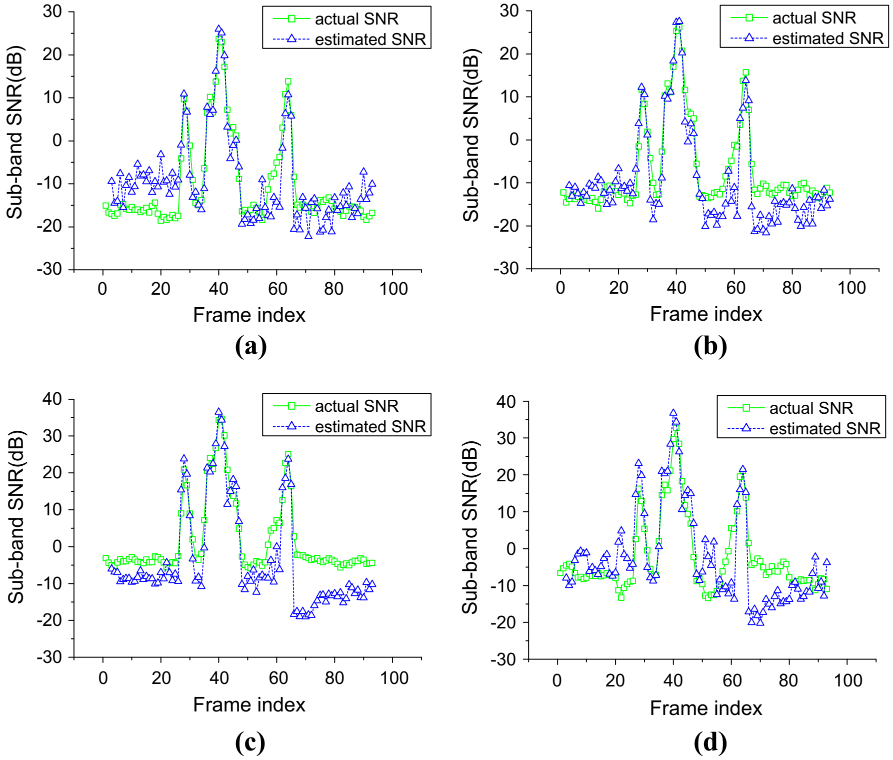


Fig. 4 SNR estimation of the 12th sub-band, **a** white, **b** pink, **c** tank, **d** speech babble

the frequency domain with time. It can not only reflect the time-variant characteristics of speech but also directly show the residual of music noise. Figure 5c is the speech spectrum of traditional spectral subtraction, and the speech spectrum shows that there is residual music noise. Because the spectral subtraction algorithm is aimed at all frequency points, the SNR varies greatly. This is why the single channel spectral subtraction will produce musical noise. However, for multi-channel spectral subtraction (Fig. 5d), the time-domain signal is decomposed into several sub-bands, and the spectral subtraction algorithm is carried out in each sub-band so that the music noise is better suppressed. In Fig. 5e, the basic Wiener filtering algorithm shows that the performance of this algorithm is deteriorated. This is because the algorithm is limited to a speech endpoint detection algorithm, and the error in the speech endpoint detection algorithm will lead to the speech frames being wrongly judged as the noise frame is being processed. Comparatively, the improved Wiener filtering algorithm, which uses the robust speech endpoint detection algorithm, is better. In addition, the patent algorithm (Fig. 5g) only works on the effect of multi-channel spectral subtraction. In contrast, the proposed algorithm (Fig. 5h) shows a better noise reduction performance.

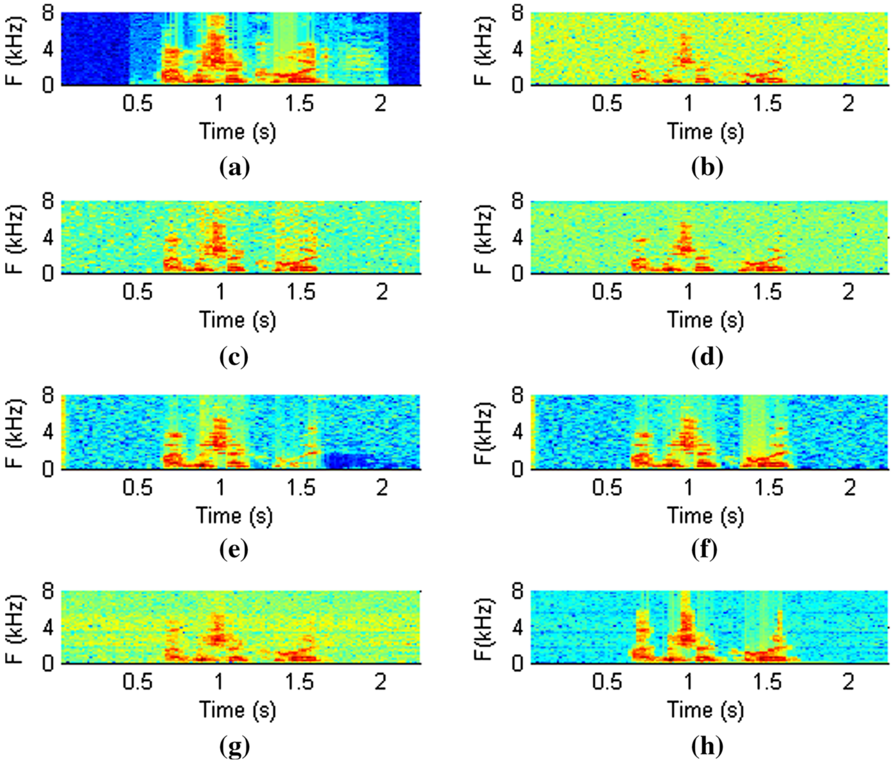


Fig. 5 Spectrograms of **a** the clean speech, **b** the noisy speech, and the **c** traditional spectral subtraction [3], **d** multi-channel spectral subtraction [18], **e** Wiener filtering [11], **f** improved Wiener filtering [23] **g** modulation frequency-based [12], and **h** proposed methods

3.2.2 Objective Performance Comparison

1. Evaluation of the log spectral measure and speech perception quality
   Because the basic spectral subtraction will produce music noise and the basic Wiener filtering method is limited to voice activity detection, the proposed algorithm will no longer be compared to the basic spectral subtraction and the Wiener filtering method in the subsequent objective index test. The experiment compared the log spectral distance (LSD) and speech perception quality (PESQ) of the improved Wiener filtering method (IWF), the sub-band spectral subtraction (MSS), modulation frequency-based (MFB), and the proposed algorithm proposed in this paper (PROP) for four types of noise when the input signal-to-noise ratio (SNR) was 0, 5 and 10 dB. The experimental results are shown in Table 3. In the white noise environment, the PESQ improvement in the algorithm proposed in this paper increased with the input signal-to-noise ratio; in the 10 dB white noise environment, its distance improved performance is similar to that of the logarithmic spectrum. The perceptual evaluation of speech quality (PESQ) improvement exhibited the following order: improved Wiener filtering method $>$ sub-band spectral subtrac-

Table 3 LSD and PESQ improvement

Noise	Input SNR	Input LSD	Input PESQ	MFB		IWF		MSS		PROP	
				Δ LSD	Δ PESQ	Δ LSD	Δ PESQ	Δ LSD	Δ PESQ	Δ LSD	Δ PESQ
White	0	20.24	1.05	↓11.74	↑0.31	↓15.66	↑1.21	↓9.76	↑0.37	↓14.73	↑0.23
	5	15.68	1.42	↓10.1	↑0.39	↓11.54	↑1.13	↓8.98	↑0.56	↓11.62	↑0.40
	10	11.30	1.80	↓7.37	↑0.43	↓7.54	↑0.77	↓7.51	↑0.51	↓7.49	↑0.47
Pink	0	16.23	1.21	↓9.48	↑0.29	↓12.19	↑0.88	↓8.16	↑0.16	↓12.09	↑0.06
	5	11.94	1.62	↓7.12	↑0.27	↓8.18	↑0.73	↓6.84	↑0.32	↓8.19	↑0.37
	10	8.03	1.99	↓4.34	↑0.26	↓4.52	↑0.61	↓4.64	↑0.39	↓4.56	↑0.37
Tank	0	9.89	1.85	↓4.71	↑0.05	↓6.14	↑0.49	↓4.51	↑0.26	↓5.98	↑0.02
	5	6.97	2.15	↓2.67	↑0.13	↓3.49	↑0.57	↓2.91	↑0.24	↓3.39	↑0.08
	10	4.81	2.42	↓1.17	↑0.09	↓1.73	↑0.18	↓1.67	↑0.17	↓1.44	↑0.12
Speech babble	0	10.95	1.49	↓3.83	0.00	↓4.80	↑0.01	↓4.13	↑0.33	↓6.40	↓0.11
	5	8.30	1.83	↓3.17	↓0.02	↓3.44	↑0.12	↓3.24	↑0.26	↓4.52	↓0.12
	10	6.05	2.13	↓2.15	↓0.02	↓2.68	↑0.21	↓2.20	↑0.2	↓2.68	↑0.18

tion > algorithm proposed in this paper > system depth method. Particularly for the algorithm proposed in this paper, the PESQ decreased by 40% compared with the improved spectral subtraction, fell by 7.8% compared with the sub-band spectral subtraction, and improved by 9.3% compared with system depth method. Under the speech babble and tank noise environments, the performances of the algorithm proposed in this paper and the system depth methods are poor, with an average PESQ improvement of no more than 0.2. The sub-band SNR estimation experiments show that the algorithm proposed in this paper has better performance under the four noise types; thus, the linear proportional attenuation model used in this paper needs to be further improved.

2. SNR performance comparison
   The output signal-to-noise ratio is an important index to evaluate the noise reduction algorithm in hearing aids. The Chinese Pure voice data were recorded in a silent room and overlaid with NoiseX-92 Library's white noise, pink noise, tank noise, and babble speech noise in the software simulation platform, which was then used as the noisy speech signal. Figure 6 gives the enhancement of the signal-to-noise ratio when 10 dB noisy speech was subjected to sub-band spectral subtraction, the improved Wiener filtering method, the system depth method and the algorithm proposed in this paper. It can be seen from the test results of Fig. 6 that in the four types of noise, the improvement in performance of the output signal-to-noise ratio exhibited the following order: improved Wiener filtering method > sub-band spectral subtraction > the algorithm proposed in this paper > system depth method. In the 10 dB white noise environment, the output SNR of the algorithm proposed in this paper decreased by 1.4 dB compared with the improved Wiener filtering method, decreased by 1.1 dB compared with the sub-band spectral subtraction method, and improved by 1 dB compared with the system depth method.
3. The real scenario experiment
   The purpose of real scenario experiment was to investigate whether the proposed method can get the good performance without AEC (acoustic echo cancelation) technology. In the previous simulation experiment, the test object is generated by the software, the noise is treated as additive noise, and the pure speech is simply overlaid with the noise to test. In practice, the existence of echo and reverberation should be considered. Therefore, the establishment of a real sense of hearing is needed in the quiet laboratory scene, and thus, real voice data were recorded using auditory scene settings as shown in Fig. 7. The left anterior and right front speakers playback noise signals to simulate the background noise, and the rear speakers play the pure speech signal to substitute for the speaker. The three speakers are placed symmetrically, and a microphone is placed in the center to gather the signal. The sound pressure level is calibrated using a TES-52A sound pressure meter. Figure 8 shows the improvement in the SNR after the noisy speech signal recorded in this real-world environment was treated by the four algorithms. It can be seen from Fig. 8 that for the four types of noise, the improved performance of the output SNR exhibited the following order: improved Wiener filtering method > sub-band spectral subtraction > algorithm proposed in this paper > system depth method. In the 10 dB white noise environment, the output SNR of the proposed algorithm is decreased by 1.9 dB compared with the improved Wiener filtering

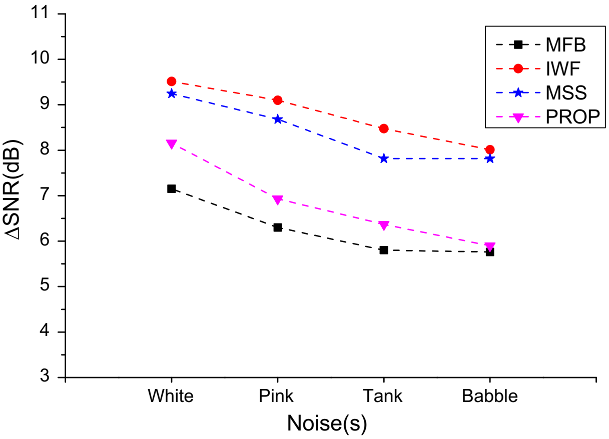


Fig. 6 SNR improvement in the simulated environment

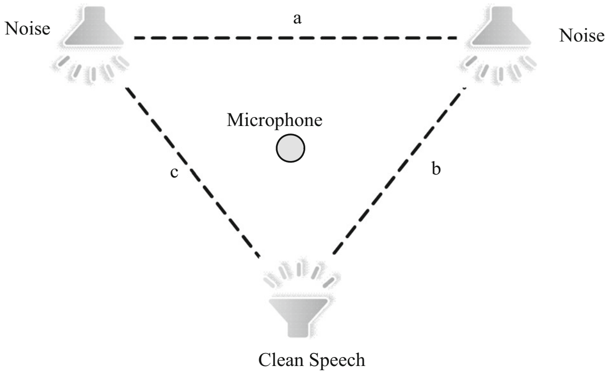


Fig. 7 Real scenario setup

method, decreased by 1.6 dB compared with the sub-band spectral subtraction method, and improved by 3.1 dB compared with the system depth method. In addition, due to the existence of echo and reverberation, this experiment also shows that the improvement in the SNR with each of the 4 algorithms decreased. Compared with the simulation environment, the improved Wiener filtering method fell by 7.9%, the sub-band spectral subtraction decreased by 12.5%, and the system depth method decreased by 12.8%, while the proposed algorithm only fell by 1.5%.

3.2.3 Subjective Performance Comparison

Sentence-pair listening test [36] was used in this subjective evaluation. The experiment included four types of noises (white, pink, tank, and babble) which were from the NoiseX-92 dataset. And the input SNR was set 10 dB. Six speech files (three male and three female) came from the TIMIT speech dataset. Comparison methods included

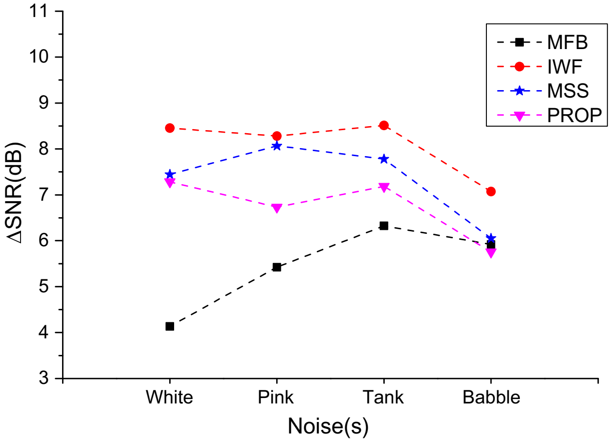


Fig. 8 SNR improvement in the real scenario

MFB, IWF, MSS, and PROP, so 24 groups enhanced speech (4*1*6) was obtained. And twelve normal hearing listeners evaluated each group enhanced speech in random order. For each group enhanced speech, listeners need to order these sentences. The scoring criterion was: +1.5 was awarded to 1st method, +1 was awarded to 2nd method, +0.5 was awarded to 3rd method, and +0 was awarded to the last. If 1st and 2nd were hard to distinguish, each method of two was awarded to +1. Also if 2nd and 3rd were hard to distinguish, each method of two was awarded to +0.5, and so on. The mean score in this subjective evaluation experiment is shown in Fig. 9. In general, the IWF enhanced speech had less residual noise and distortion. MSS and PROP methods had equivalent performance on residual noise and distortion. The MFB method exits much residual noise.

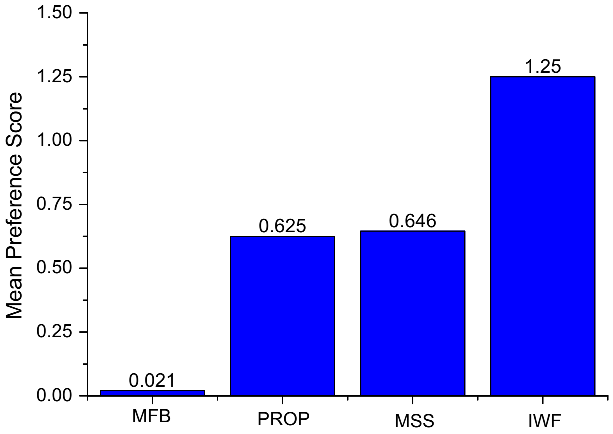


Fig. 9 Comparisons of mean preference score among algorithms

3.2.4 Real-Time Performance Comparison

Figure 10 shows the average processing delay of the proposed algorithm (PROP), modulation depth method [12], improved Wiener filtering method [10], and sub-band spectrum subtraction [18]. The times of processed noisy speech are 0.55, 1.1, 2.26 s, the frame size is 256, and the frame shift is 50%. Although an analysis filter is needed for the multi-band algorithm to divide the signal into a number of sub-bands, filter decomposition is the basic operation of hearing aid algorithms, and its delay is not included in the experiment's performance comparison. Thus, the statistical processing delay is measured from the signal passing through the decomposition filter to reaching the synthesis filter. In Fig. 10, the delays of the sub-band spectrum subtraction method and improved Wiener filtering method are significantly larger than those of the proposed method and modulation depth method. The processing delay performance of the proposed algorithm is decreased by 51% compared with the modulation depth method but improved by 60.6% compared with the sub-band spectrum subtraction method and by 40.7% compared with the improved Wiener filtering method. This is because the two models must use the fast Fourier transform to analyze the signal in the frequency domain and ultimately restore to signal to the time domain by the inverse fast Fourier transform; thus, the amount of calculation will be increased significantly. And based on the above analysis, this method can be implemented at about 25% calculation cost of real and imaginary modulation spectral subtraction method. However, in the modulation depth method proposed in [2], the amount of gain is determined by the signal envelope, the processing delay is reduced greatly and the delay reaches a minimum by avoiding the FFT and IFFT computation. The processing delay of the proposed algorithm is longer than that of the modulation depth method but much lower than those of the spectral subtraction method and Wiener filtering method. This is because the estimate of the sub-band SNR in the proposed algorithm is also performed in the frequency domain, the FFT calculations are only used to estimate the sub-band SNR,

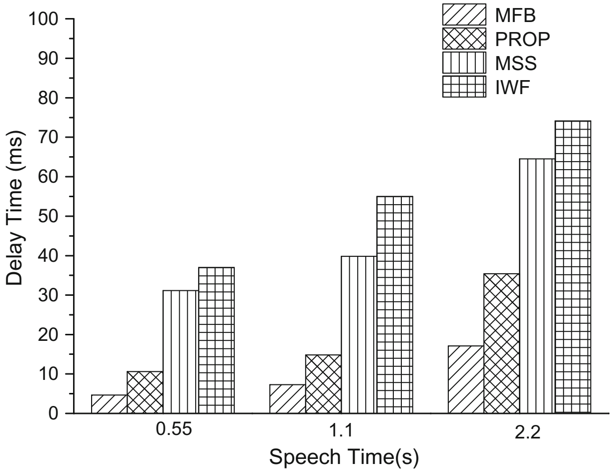


Fig. 10 Comparisons among algorithms of the time delay performance

and the cross-correlation function and mean square value calculations involved are not highly time-consuming. Additionally, the core advantages of the method is retaining the algorithm features of the literature [25] and using the time-domain gain function instead of the inverse fast Fourier transform (IFFT) to restore the time-domain signal. This is the fundamental reason for the decrease by half processing time compared with the spectral subtraction method and Wiener filtering method.

4 Conclusions

A multi-channel digital noise reduction algorithm based on sub-band SNR estimation is proposed. The algorithm has satisfied performance in noise suppression and maintains a low time complexity. The algorithm estimates the speech signal based on maximum a posteriori probability theory, and the adjacent frame information in the form of the joint probability density function is introduced to the maximum a posteriori probability model. The estimated value of the speech signal is derived by calculating the autocorrelation matrix of two adjacent frames, wherein the noise estimate uses the minimum noise spectrum to track the noise. To verify the performance of the algorithm, many experiments are designed, and various algorithms are compared by using LSD, PESQ, SNR, and processing delay indicators. Under the environment of 10 dB white noise, the improvement in the spectrum distance exhibits the following order: modulation depth method > improved Wiener filtering method > sub-band spectral subtraction method > proposed method. The improvement in speech perceptual quality exhibits the following order: improved Wiener filtering > sub-band spectral subtraction method > proposed method > modulation depth method. The order of improvement in the output SNR in the simulation environment is as follows: improved Wiener filtering method > sub-band spectral subtraction method > proposed method > modulation depth method. The order of improvement in the output SNR in the real scenario is as follows: improved Wiener filtering method > sub-band spectral subtraction method > proposed method > modulation depth method. The order of the delay size of the four algorithms is as follows: modulation depth method < proposed method < sub-band spectral subtraction method < improved Wiener filtering method. Therefore, considering the speech distortion, perceptual quality and processing delay performance, the proposed method is an optional program for low-power devices such as digital hearing aids.

Of course, there are some limitations for the proposed algorithm. This model is under the assumption of a Gaussian model, so it is not suit for nonstationary noise scene, such as babble noise. The babble noise is one of the most important conditions to improve. The gain functions are set the same in the four noisy environments, resulting in the noise reduction of the algorithm not being suited to the tank and speech babble background noises. Thus, auditory scene classifier will be the direction of the follow-up research on the algorithm. And the parameters of gain function are adjusted according to the different noise scenes.

Acknowledgements The work was supported by the National Natural Science Foundation of China under Grant Nos. 61673108 and 61375028, China Postdoctoral Science Foundation funded Project under Grant No. 2016M601696, Qing Lan Project of Jiangsu Province, Six Talent Peaks Project in Jiangsu Province under Grant No. 2016-DZXX-023, Jiangsu Planned Projects for Postdoctoral Research Funds under Grant No. 1601011B, the High-level Scientific Research Foundation of NJIT under Grant No. YKJ201526 and the open fund of Guangdong lighting and audio video engineering research center under Grant Nos. KF201601 and KF201602. The authors would like to thank the reviewers for their valuable suggestions and comments.

Compliance with Ethical Standards

Conflict of interest All authors declare that they have no conflict of interest.

References

1. B. Acar, M.F. Yurekli, M.A. Babademez, H. Karabulut, R.M. Karasen, Effects of hearing aids on cognitive functions and depressive signs in elderly people. Arch. Gerontol. Geriatr. **52**(3), 250–252 (2011). doi:10.1016/j.archger.2010.04.013
2. R. Aggarwal, J.K. Singh, V.K. Gupta, S. Rathore, M. Tiwari, A. Khare, Noise reduction of speech signal using wavelet transform with modified universal threshold. Int. J. Comput. Appl. **20**(5), 14–19 (2011)
3. S.F. Boll, Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. Acoust. Speech Signal Process. **27**(2), 113–120 (1979)
4. K.S. Chong, B.H. Gwee, J.S. Chang, A 16-channel low-power nonuniform spaced filter bank core for digital hearing aids. IEEE Trans. Circuits Syst. II Express Br. **53**(9), 853–857 (2006)
5. R. Chou, T. Dana, C. Bougatsos, C. Fleming, T. Beil, Screening adults aged 50 years or older for hearing loss: a review of the evidence for the US preventive services task force. Ann. Int. Med. **154**(5), 347–355 (2011)
6. K. Chung, Challenges and recent developments in hearing aids part I. Speech understanding in noise, microphone technologies and noise reduction algorithms. Trends Amplif. **8**(3), 83–124 (2004)
7. I. Cohen, Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging. IEEE Trans. Speech Audio Process. **11**(5), 466–475 (2003)
8. K.L. Cummins, K.E. Hecox, M.J. Williamson, Adaptive, programmable signal processing hearing aid, in EP (1989)
9. B. Edwards, The future of hearing aid technology. Trends Amplif. **11**(1), 31–46 (2007)
10. M.A.A. El-Fattah, M.I. Dessouky, A.M. Abbas, S.M. Diab, E.-S.M. El-Rabaie, W. Al-Nuaimy, S.A. Alshebeili, F.E.A. El-Samie, Speech enhancement with an adaptive Wiener filter. Int. J. Speech Technol. **17**(1), 53–64 (2014)
11. Y. Ephraim, D. Malah, Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. IEEE Trans. Acoust. Speech Signal Process. **33**(2), 443–445 (1985)
12. X. Fang, M.J. Nilsson, Noise reduction apparatus and method, in US (2004)
13. B. Gopinath, J. Schneider, D. Hartley, E. Teber, C.M. McMahon, S.R. Leeder, P. Mitchell, Incidence and predictors of hearing aid use and ownership among older adults with hearing loss. Ann. Epidemiol. **21**(7), 497–506 (2011)
14. Y. Gui, H.K. Kwan, Adaptive subband Wiener filtering for speech enhancement using critical-band gammatone filterbank. Midwest Symp. Circuits Syst. **731**, 732–735 (2005)
15. A. Hogan, K. O’Loughlin, P. Miller, H. Kendig, The health impact of a hearing disability on older people in Australia. J. Aging Health **21**(8), 1098–1111 (2009)
16. M.T. Islam, C. Shahnaz, W.-P. Zhu, M.O. Ahmad, Rayleigh modeling of teager energy operated perceptual wavelet packet coefficients for enhancing noisy speech. Speech Commun. **86**, 64–74 (2017)
17. Y. Jiang, R. Liu, Binaural deep neural network for robust speech enhancement, in *2014 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)* (2014), pp. 692–695
18. S. Kamath, P. Loizou, A multi-band spectral subtraction method for enhancing speech corrupted by colored noise, in *IEEE International Conference on Acoustics Speech and Signal Processing* (Citeseer, 2002), pp. 4164–4164
19. H. Katahira, N. Ono, S. Miyabe, T. Yamada, S. Makino, Nonlinear speech enhancement by virtual increase of channels and maximum SNR beamformer. EURASIP J. Adv. Signal Process. **2016**, 11 (2016)

20. S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* (1993), pp. 342–343. doi:10.2307/1269750
21. M.C. Killion, P.A. Niquette, What can the pure-tone audiogram tell us about a patient's SNR loss? Hear. J. **53**(3), 46–48 (2000)
22. A. Laplante-Lévesque, L. Hickson, L. Worrall, Rehabilitation of older adults with hearing impairment: a critical review. J. Aging Health **22**, 143–153 (2010)
23. J. Lei, J. Wang, Z. Yang, Robust voice activity detection based on spectral entropy and two-stage mel-warped Wiener filtering (IEEE, 2008) pp. 306–309
24. F.R. Lin, R. Thorpe, S. Gordon-Salant, L. Ferrucci, Hearing loss prevalence and risk factors among older adults in the United States. J. Gerontol. Ser. A: Biol. Sci. Med. Sci. **66**(5), 582–590 (2011)
25. R. Martin, Noise power spectral density estimation based on optimal smoothing and minimum statistics. IEEE Trans Speech Audio Process. **9**(5), 504–512 (2001)
26. S. Mavaddaty, S.M. Ahadi, S. Seyedin, Speech enhancement using sparse dictionary learning in wavelet packet transform domain. Comput. Speech Lang. **44**, 22–47 (2017)
27. M.A.B. Messaoud, A. Bouzid, Speech enhancement based on wavelet transform and improved subspace decomposition. J. Audio Eng. Soc. **63**(12), 990–1000 (2016)
28. S.-F. Ou, X.H. Zhao, MAP estimation for noisy speech enhancement based on inter-frame correlation. Acta Electron. Sin. **35**(10), 2007–2013 (2007)
29. J.W. Seok, K.S. Bae, Speech enhancement with reduction of noise components in the wavelet domain, in *1997 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1997. ICASSP-97* (IEEE, 1997), pp. 1323–1326
30. H.W. Tseng, M. Hong, Z.Q. Luo, Combining sparse NMF with deep neural network: a new classification-based approach for speech enhancement, in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2015)
31. Q. Wang, R. Liang, Z. Zou, L. Zhao, Sub-band noise reduction in multi-channel digital hearing aid. IEICE Trans. Inf. Syst. **E99.D**(1), 292–295 (2016)
32. J. Wouters, J.V. Berghe, Speech recognition in noise for cochlear implantees with a two-microphone monaural adaptive noise reduction system. Ear Hear. **22**(5), 420–430 (2001)
33. S. Wyrsch, A. Kaelin, Subband signal processing for hearing aids. In: *Proceedings of the 1999 IEEE International Symposium on Circuits and Systems, 1999. ISCAS '99*, vol. 23 (Jul 1999), pp. 29–32
34. L.C. Yang, Q.Y. Tao, W.W. Hong, A GSC algorithm based on null spectral subtraction for dual small microphone array speech enhancement. Zhejiang Daxue Xuebao (Gongxue Ban)/J. Zhejiang Univ. (Eng. Sci. Ed.) **47**(8), 1493–1499 (2013)
35. P. Yotam, R. Boaz, Objective performance analysis of spherical microphone arrays for speech enhancement in rooms. J. Acoust. Soc. Am. **132**(3), 1473–1481 (2012)
36. Y. Zhang, Y. Zhao, Real and imaginary modulation spectral subtraction for speech enhancement. Speech Commun. **55**(4), 509–522 (2013)