33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Noise Reduction for Binaural Hearing Aids Using Unsupervised Diffuse Noise Estimator

Youna Ji, Youngcheol Park, Dongwook Kim, and Junil Sohn

Abstract— **In this paper a new noise reduction algorithm for binaural hearing aids is proposed. This algorithm is capable of suppressing both nonstationary diffuse noise and unknown directional interferences without distorting the directional cues. For the estimation of the diffuse noise power spectral density (PSD), we utilize the eigenstructure of the 2x2 input covariance matrix, together with a compensation process for preventing the underestimation at low frequencies. The interference PSD is estimated from the target cancelled input signals through the signal prediction. Effectiveness of the proposed algorithm was confirmed according to the computer simulations in terms of noise reduction and binaural cue preservation.**

I. Introduction

For many years, speech enhancement for hearing aids has been as issue of great concern to the hearing aid manufacturers. Main goal of speech enhancement is suppressing the noise as much as possible without introducing speech distortion. Many algorithms have been developed in an effort to achieve this goal. A large number of previous techniques use voice activity detector (VAD) to estimate noise power spectral density (PSD). However VAD system for the binaural hearing aids has to be sophisticated enough to distinguish between directional interference and target speech. Also, most of noise estimation algorithms developed under the uncorrelation assumption suffer from underestimation of noise PSD at low frequencies, because, in real environment, the diffuse noises such as cafeteria noise or reverberation, show high coherence at low frequencies [1].

For binaural hearing aid users maintaining directivity of the sound is important. In certain situation, such as traffic, incorrect localization of sounds could even endanger the users. Human auditory system recognizes direction of sound based on spatial cues. In particular, azimuth cues such as interaural time difference (ITD) and interaural level difference (ILD) are known to play important roles for the perception sound direction [2]. Thus, even after the noise suppression process, these binaural cues have to be preserved to avoid directivity loss. The state of art technique uses multi-channel microphone to reduce the noises and preserve binaural cues. However, use of multiple microphones need more power consumption. Also due to the size and weight limits of hearing aids, arrays with more than two microphones cannot be accommodated.

Youna Ji and Youngcheol Park are with Computer & Telecomm. Eng. Division, Yonsei University, 1 Yonseidae-gil Wonju, Kwangwon-do, Korea (Phone:+82-33-760-2744, e-mail : young00@yonsei.ac.kr)

Dongwook Kim and Junil Sohn are with Samsung Advanced Institute of Technology (SAIT), Suwon, Korea. (e-mail : ji.sohn@samsung.com)

In this paper, we propose a new noise reduction algorithm for binaural hearing aids with a single microphone on each side. The proposed algorithm is capable of suppressing both nonstationary diffuse noise and directional interferences with minimum loss of directional cues such as ILD and ITD. For the estimation of diffuse noise, we propose an unsupervised noise power spectral density (PSD) estimator, which is based on the eigenstructure of the (2x2) input spatial covariance matrix. Since the proposed method can estimate the noise PSD even during the presence of speech, it doesn't require VAD. Also, using a compensation process, the proposed estimator can prevent underestimation of PSD at low frequencies.

For the estimation of interference, the target component presumed to be known is first eliminated from the input channels. Interferences are then estimated through the process of signal prediction. In this process, we propose three different approaches by differing target components for the PSD estimation. Finally, using the estimated diffuse noise and interference PSDs, appropriate channel gains are calculated. The main advantage of this approach is that it is robust and effective regardless of the direction or number of interference.

The remainder of this paper is organized as follows: Section II presents the proposed algorithm. The simulation results will be provided in Section III. Finally conclusions are drawn in Section IV.

II. Process of Binaural Noise Reduction Algorithm

The proposed binaural noise reduction algorithm is depicted in Fig.1.

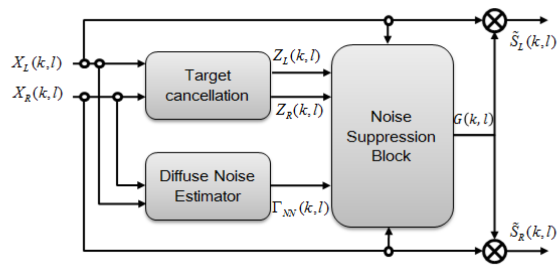

Fig 1. Block diagram of the proposed binaural noise reduction algorithm.

When the binaural noisy input signals enter the system, the target signal presumably known is cancelled in the target cancellation block. As a result, it produces outputs as sum of diffuse and interference signals. On the other hand, the diffuse noise PSD is independently estimated from the input noisy signal.

The target-cancelled noisy signals are delivered to the noise suppression block, in which channel gains for the noise reduction are calculated. We use 3 different approaches to the gain calculation by differing the target component for PSD estimation.

A. The Binaural Noisy Input Signals

In complex noise environment, the binaural noisy input signals $x_L(\mathrm{t})$, and $x_R(\mathrm{t})$, are modeled as

$$x_i(\mathrm{t}) = s(t) * h_i(t) + v(t) * g_i(t) + n_i(t),\ i = L, R \quad (1)$$

where $s(t)$ is noise free target signal and $v(t)$ is interference which comes from certain direction and $n_i(t)$ is diffuse noise. Also, $h_i(t)$ and $g_i(t)$ are the Head Related Impulse Responses (HRIRs) related to the directions of target and interference, respectively. Although we assume that the direction of target signal is known, there is no a priori information about interference direction. The interference can be anywhere and can be composed of multiple sources. The diffuse noise signal is propagating from all directions with equal amplitude and a random phase e.g. bubble noise in cafeteria and reverberation [2].

Input signals can be analyzed in frequency domain using FFT. Then, Eqs. (1) can be rewritten as

$$X_i(k,l) = S(k,l)H_i(k,l) + V(k,l)G_i(k,l) + N_i(k,l) \quad (2)$$

where k and l are frequency and frame indices, respectively.

B. The Diffuse Noise Estimation

For simplicity, we assume that each frequency bin contains only one directional signal component and uncorrelated diffuse noise. The directional signal component in each frequency bin is from either target or interference. Then, noisy input signal in frequency domain can be simplified as

$$X_i(k,l) = D(k,l)C_i(k,l) + N_i(k,l) \quad (3)$$

where $D(k,l)$ is target or interference component and $C_i(k,l)$ is head related transfer function (HRTF) corresponding to its direction. Now, the covariance matrix of noisy input signal is defined as

$$\mathbf{R} = \begin{bmatrix} |C_L|^2\Gamma_{DD} + \Gamma_{NN} & C_L C_R^*\Gamma_{DD} + \Gamma_{NN}^{LR} \\ C_R C_L^*\Gamma_{DD} + \Gamma_{NN}^{RL} & |C_R|^2\Gamma_{DD} + \Gamma_{NN} \end{bmatrix} \quad (4)$$

where $\Gamma_{DD} = E\{|D|^2\}$ and $\Gamma_{NN} = E\{|N_L|^2\} = E\{|N_R|^2\}$. Also, $\Gamma_{NN}^{LR} = E\{N_L N_R^*\}$ denotes cross PSD between the noise components in the left and right channels. For simplicity, we omit frequency and frame indices. If we assume that left and right channel of diffuse noise are uncorrelated, the cross PSD Γ_{NN}^{LR} becomes 0. Then, eigenvalues of the covariance matrix can be obtained from the characteristic equation, as given by

$$\lambda_{1.2} = \frac{\left((|C_L|^2+|C_R|^2)\Gamma_{DD}+2\Gamma_{NN}\right)\pm\sqrt{\Delta}}{2} \quad (5)$$

where $\Delta = \left((|C_L|^2 + |C_R|^2)\Gamma_{DD}\right)^2$. Thus, it is straightforward to see that the smaller eigenvalue with minus sign, i.e., λ_2, is identical to the diffuse noise PSD Γ_{NN}.

But the estimated noise PSD in above equation often suffers from underestimation, especially at low frequencies below 500Hz, because diffuse noise in real environment has high coherence at low frequencies [3]. The coherence function, on the other hand, can be modeled using sinc function as $\Psi = \sin\left(\frac{2\pi f d_{LR}}{c}\right)$ [3]. Using this coherence model, the cross PSD of the left and right noise signals can be approximated as $\Gamma_{NN}^{LR} \approx \Psi\Gamma_{NN}$. In addition, ILDs at low frequencies are negligible [4]. Based on these, we have

$$\Delta \approx \left((|C_L|^2 + |C_R|^2)\Gamma_{DD} + 2\Psi\Gamma_{NN}\right)^2. \quad (6)$$

Now, by rearranging the terms, diffuse noise PSDs at low frequencies are approximated as

$$\Gamma_{NN} \approx \frac{\lambda_2}{1-\Psi}. \quad (7)$$

C. Target Cancellation Block

The purpose of target cancellation block is to obtain noise plus interference only signal from the noisy input. Since we assumed that the target direction is known *a priori*, the target-cancelled signals can be conveniently obtained as

$$Z_L = X_L - W_R^T X_R = a_L V + N_L' \quad (8)$$
$$Z_R = X_R - W_L^T X_L = a_R V + N_R' \quad (9)$$

where $W_R^T = \frac{H_L^* H_R}{|H_L|^2}$ and $W_L^T = \frac{H_R^* H_L}{|H_R|^2}$. Later, the interference PSD is estimated from the target-cancelled signals Z_L and Z_R. The main advantage of this approach is that there is no restriction on the number, location and content of the interference [5].

D. Noise Suppression Block

In the noise suppression block, the interference PSD is estimated, and appropriate channel gains are calculated to suppress both the diffuse noise and interference. In all cases, the gain factors for noise suppression are shared in both left and right channel to preserve spatial cue [5].

In this paper, the target-canceled signal is used to estimate signal components in the input channels. The target or interference signal is found from an optimization problem: $min_w \|X_i - W_i^{N*} Z_i\|_2$, where $\|.\|_2$ denotes 2-norm. The target or interference component can be accurately estimated from the above optimization problem, by assuming that all signals are uncorrelated. However, the diffuse noise component in Z_i, which can be significant in real-life situations, hinders estimation of the signal components.

To handle this issue, we suggest three different methods for constructing the noise suppression block. Details of each approach are presented below.

Method 1: Fig. 2 shows a schematic diagram of *Method 1*. In this method, the diffuse noise components N_i comprised in both X_i and Z_i are suppressed (left dashed box in Fig. 2), prior to the signal estimation. Then, the target signal is estimated using the remaining signals (right dashed box in Fig. 2). Finally, the gain factors are calculated using the estimated target signals.

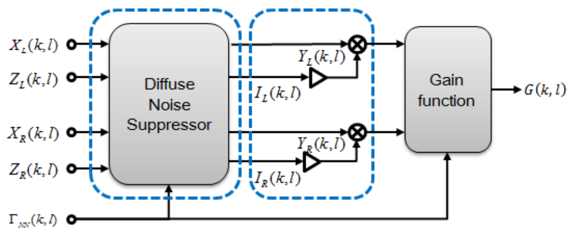

Fig. 2 Noise suppression block with *Method 1*.

The diffuse noise suppression is conducted as

$$Y_i = X_i \cdot G_i^b, \quad I_i = Z_i \cdot G_i^c \tag{10}$$

where $G_i^b = \frac{\Gamma_{YY}^i}{\Gamma_{XX}^i}$ and $G_i^c = \frac{\Gamma_{II}^i}{\Gamma_{ZZ}^i}$ are the gain factors for suppressing the diffuse noise. Parameters for the gain factors are estimated as

$$\Gamma_{YY}^i = \Gamma_{XX}^i - \hat{\Gamma}_{NN}, \tag{11}$$
$$\Gamma_{II}^L = \Gamma_{ZZ}^L - (1 + |W_R^T|^2)\hat{\Gamma}_{NN} \tag{12}$$
$$\Gamma_{II}^R = \Gamma_{ZZ}^R - (1 + |W_L^T|^2)\hat{\Gamma}_{NN} \tag{13}$$

where $\hat{\Gamma}_{NN}$ denotes the estimate of diffuse noise PSD. After the diffuse noise suppression, the target estimation is conducted using Y_i and I_i. For recursive estimation, we choose the Normalized Least Mean Square (NLMS) algorithm, which is summarized as

$$\hat{S}_i = Y_i - A_i^l \cdot I_i, \tag{14}$$
$$A_i^{l+1} = A_i^l + \mu \frac{I_i^*}{\hat{\Gamma}_{II}^i} \cdot \hat{S}_i, \tag{15}$$

where μ is step-size, l denotes frame index and $\hat{S}_i$ is the estimate of the target signal in the channel i.

Finally, using the target estimate together with the PSD of the input noisy signal, the channel gains are calculated. To preserve the spatial cues, a gain function is shared both for left and right channel. The gain function is calculated as

$$Gain_i = \frac{|\hat{S}_i|^2}{\Gamma_{XX}^i}, \tag{16}$$
$$G^{inter} = \sqrt{Gain_L \cdot Gain_R}\,. \tag{17}$$

The gain G^{inter} is to suppress the interference. Thus, for the reduction of diffuse noise, we have to generate another attenuation factor for the suppression of diffuse noise. For this end, the modified Wiener filter based on *a priori* SNR can be utilized. *A priori* SNR is calculated using the decision-directed method which is known to be effective for reducing musical noise [7]. The, we have

$$G^{diff} = \frac{\xi}{1+\xi}, \tag{18}$$

where ξ is a priori SNR for the binaural input given by $\xi = \left(|\hat{S}_L|^2 + |\hat{S}_R|^2\right)/2\hat{\Gamma}_{NN}$.

The final gain of noise suppression block is obtained by combining the interference suppression gain in Eq. (17) and the diffuse suppression gain in Eq. (18) as

$$G^{final} = \sqrt{G^{inter} \cdot G^{diff}}. \tag{19}$$

Method 2: In the previous method, we suppress diffuse noise before the target estimation. However, for simplicity, we can conduct the target estimation directly using the input and target-cancelled signals without the diffuse noise suppression. In this method, the NLMS algorithm for the target signal estimation is modified as

$$\tilde{S}_i = X_i - B_i^l \cdot Z_i, \tag{20}$$
$$B_i^{l+1} = B_i^l + \mu \frac{Z_i^*}{\hat{\Gamma}_{ZZ}^i} \cdot \tilde{S}_i. \tag{21}$$

Now, the gain function is calculated using Eqs. (17), (18), and (19) using $\tilde{S}_i$ instead of $\hat{S}_i$.

Method 3: Through the process of signal estimation, we have obtained diffuse noise PSD, and target and interference estimates. Based on the estimated signal components, we can directly apply the Wiener filter to the noisy input signals. Thus, in *Method 3*, the gain function for suppressing the interference is determined as

$$G_i^{inter} = \frac{\xi_i}{1+\xi_i} \tag{22}$$

where ξ_i is a priori SNR given by $\xi_i = |\hat{S}_i|^2/\hat{\Gamma}_{II}^i$, and it is updated using the decision-directed scheme. Thus, the second dashed block in Fig 2 is unnecessary, and the gain function for the diffuse noise suppression is also determined using Wiener filter.

The gain functions for all 3 methods are summarized in Table 1. The enhanced binaural outputs are obtained by multiplying the final gains to the left and right input signals, respectively, as $\hat{S}_i = X_i \cdot G^{final}$. Finally, the enhanced signals are transformed to the time-domain signal through IFFT.

TABLE I. GAIN FUNCTIONS FOR THE INTERFERENCE SUPPRESSION.

	Gain for left or right channel
Method 1	$G_i^{inter} = \frac{|Y_i - W_i^n \cdot I_i|^2}{\Gamma_{XX}^i}$
Method 2	$G_i^{inter} = \frac{|X_i - W_i^n \cdot Z_i|^2}{\Gamma_{XX}^i}$
Method 3	$G_i^{inter} = \frac{\xi_i}{1+\xi_i}, \xi_i = \frac{|S_i|^2}{\Gamma_{II}^i}$
Diffuse Noise Gain	$G^{diff} = \frac{\xi}{1+\xi}, \xi = \frac{(|S_L|^2 + |S_R|^2)}{2\Gamma_{NN}}$

III. Simulation Results

In this section performances of the proposed algorithms are evaluated and compared. The noisy input signal segmented into sub frames of 23.2ms with 50% overlap at 22.05kHz sampling rate. A sine window is applied to each frame of 512 samples.

In the first selected complex acoustical scenario, a female target voice was in front of hearing aid user and a male interference voice was at 120° to the right. Cafeteria noise was added to the input channels at 3dB and 0dB SNRs for left and right channels, respectively. In the second scenario, target and interference were located at 90° and 210°, respectively. Cafeteria noise was also added at -3dB and 3dB SNRs.

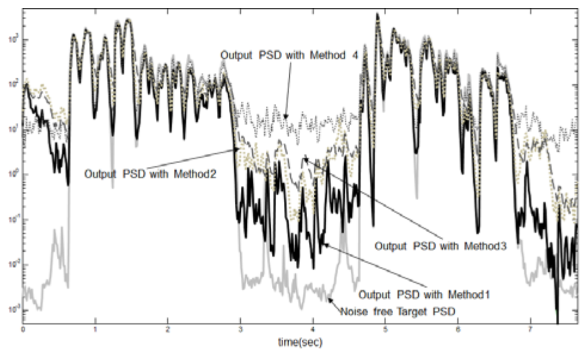

Fig. 3 Sum of enhanced target PSDs.

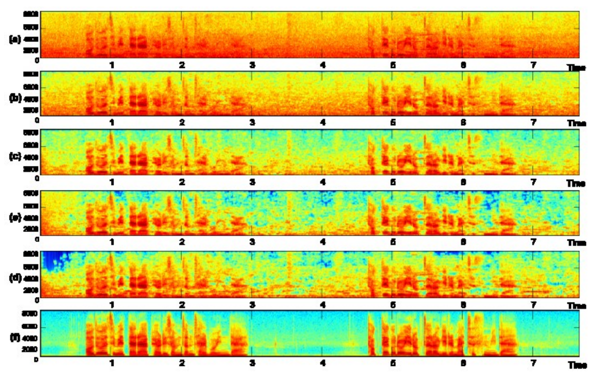

Fig 4. Spectrograms of (a) noisy input, outputs of (b) Method 4 in [5]. (c) Method 1, (d) Method 2, (e) Method 3, (f) clean speech Spectrogram

Fig. 3 shows output PSDs obtained in Scenario 1. The plots were obtained summing the enhanced target PSD in each frame. For comparison, an advanced binaural noise reduction algorithm in [5] was compared. This algorithm estimates interference using equalization - cancellation (EC) process, and then apply Wiener filter to obtain an enhanced target signal. The method in [5] is referred to as *Method 4*. First of all, it can be seen that the proposed methods outperform the previous method (*Method 4*). *Method 4* cannot effectively reduce the diffuse noise for all time period. Among the proposed methods, *Method 1* shows the best performance in terms of attenuating noises in non-speech region. In addition speech quality of *Method 4* was not as good as the proposed methods because of the unsuppressed noises.

For further comparison between the proposed methods, we measured performance parameters. For a measure of speech enhancement, frequency weighted SNR [7] and SNR improvement [5] were calculated. And for a measure of binaural cue preservation, ITD and ILD errors [5] were calculated. Results are summarized in Table 2 for Scenario 1 and Scenario 2.

TABLE 2. Results of performance parameter measures.

	Method 1		Method 2		Method 3		Method 4	
Target direction	0°	90°	0°	90°	0°	90°	0°	90°
EITD	0.05	0.07	0.03	0.06	0.1	0.11	0.03	0.07
EILD	2.5	4.2	3.50	7.0	3.7	7.4	3.4	6.9
Improvement SNR	6.1 / 7.1	9.0 / 4.0	6.5 / 7.0	8.8 / 5.3	5.3 / 5.5	7.1 / 4.6	4.9/ 6.0	8.0/ 5.2
fwSNR	4.5/ 4.1	4.4/ 5.5	5.0 / 4.6	4.4/ 3.8	4.5 / 3.8	3.5 / 3.7	4.7/ 4.3	4.4/ 5.4

All the proposed methods show small ITD and ILD errors, so that minimal loss of spatial cues is confirmed. However when target is located at directions other than front, *Method 3* showed large increase of ILD error. It was due to inaccuracy of the approximated interference PSD that controls the Wiener gain. Among the proposed methods, *Method 1* showed the best results. It showed the smallest ILD and ITD errors and the biggest improvement of SNR and highest frequency weighted SNR regardless of target direction.

IV. Conclusion

A new noise reduction algorithm for binaural hearing aids was proposed. The diffuse noise PSD was estimated in a form of eigenvalue. Three different methods were suggested to suppress noise. Simulation results showed that the proposed methods attenuate noise component effectively regardless of speech presence and direction of speech. Also, all proposed methods do not impair the spatial cues. Especially *Method 1* and *Method 2* show better performance parameters than previous technique, *Method 4*.

References

[1] R. Martin, *Noise power spectral density estimation based on optimal smoothing and minimum statistics, IEEE Trans. Acoust., Speech Signal Processing,* vol. 9, pp 504-512, July 2001.

[2] H. Abutalebi, H. Sheikhzadeh, and L. Brennana , *A hybrid subband adaptive system for speech enhancement in diffuse noise fields*. *IEEE Signal Process Lett*., vol. 11, pp. 44-47, Jan. 2004

[3] A. H. Kamkar-Parsi and M. Bouchard, *Improved noise power spectrum density estimation for binaural hearing aids operating in a diffuse noise filed environment, IEEE Trans, Audio, Speech , Lang. Process.*, vol. 17, no. 4, pp. 521-533, May 2009.

[4] V. Ralph Algazi, Carlos Avendano, and Richard O. Duda, *Elevation localization and head-related transfer function analysis at low frequencies*, J. Acoust. Soc. Am. Vol. 109, iss. 2 pp. 1110-1122, 2001.

[5] Junfeng Li, Shuichi Sakamoto, Satoshi Hongo, Masato Akagi, Yoiti Suzuki, *Two-stage binaural speech enhancement with Wiener filter for high-quality speech communication*, Speech Comm. 2010

[6] Y. Ephraim and D. Malah, *Speech enhancement using a minimum mean square error short time spectral amplitude estimator, IEEE Trans, Acoust, Speech, Signal processing* vol. ASSP-32, pp. 1109-1121, Dec 1984.

[7] Yi Hu and Philipos C. Loizou, *Evaluation of Objective Quality Measures for Speech Enhancement, IEEE Trans. Audio, Speech, Lang. Processing*, Vol. 16, No .1 Jan. 2008.