BINAURAL NOISE SUPPRESSION BASED ON AN UNBIASED ESTIMATOR OF TARGET PSD IN COMPLEX NOISE ENVIRONMENTS

Youna Ji, Young-cheol Park

Computer and Telecomm. Eng. Division Yonsei University Wonju, Korea

ABSTRACT

The conventional target power density spectrum (PSD) estimation methods based on the signal prediction inherently produce a biased target PSD because of irrelevant assumptions for the noisy environment. In this paper, an unbiased target PSD is obtained by removing the effect of diffuse noise on the prediction filter. In addition, by on-line estimation of both the noise PSD and target transfer function ratio (TFR) from the input signals, the proposed algorithm achieves robust noise suppression for an unknown target direction under a fast time-varying noisy environment. Computer simulations demonstrate the effectiveness and superiority of the proposed algorithm over the conventional methods.

Index Terms— target PSD estimation, binaural hearing aids, interference reduction

1. INTRODUCTION

By wearing hearing aids in both ears, it is significantly easier to understand speech in noisy environments. The goal of binaural noise suppression is to exploit further benefits for improving user comfort and speech intelligibility by reducing the environmental noise in a binaural signal. Real-life acoustic environments consist of various sound components, such as the desired speech signal, diffuse noise, and directional interference. Accordingly, speech enhancement algorithms have to deal with all of the noise components in a unified framework.

Two-stage binaural speech enhancement with a Wiener filter (TS-BASE/WF) [1] is a recently proposed binaural speech enhancement algorithm. In this algorithm, the interference component is first obtained after removing the target signal based on equalization and cancellation (EC) theory, and then the estimated interference component was compensated using a Wiener filter. It was proven that the algorithm could effectively deal with multiple directional noise sources and preserve binaural cues. However, this method has no consideration for diffuse noise characteristic, so that the performance degradation is inevitable in an environment where Daehee Youn

Department of Electrical & Electronic Eng. Yonsei University Seoul, Korea

diffuse noise is substantial. More recently, binaural instantaneous estimator-based noise reduction scheme (BIENR) was proposed [2]. This algorithm retrieved a target PSD in a noisy environment using a linear prediction filter model. However, this algorithm assumed that the noise field surrounding the user of a hearing aid was dominated by directional interference sources. Thus, if there is diffuse background noise with relatively significant power, the produced PSD estimates can be biased, which eventually results in degradation in the sound quality at the output of the noise suppression system.

In this paper, we propose an unbiased target PSD estimator for binaural digital hearing aids operating in a complex acoustic noisy environment. We numerically analyze the bias of the prediction filter due to the diffuse noise, and a noise suppression system based on the proposed target PSD estimator is presented. The proposed algorithm alleviates the effect of diffuse noise on the target PSD estimation by computing the minimum mean square error (MMSE) of the prediction filter using the interference component only. In this process, the transfer function ratio (TFR) of the target signal and the diffuse noise PSD are assumed to be known, and in the proposed binaural noise suppression system, those parameters are estimated on-line.

2. UNBIASED TARGET PSD ESTIMATION ALGORITHM

In an environment with complex noise, the binaural input signals can be represented in the frequency domain as

$$X_{i}(k,l) = S(k,l)H_{i}(k,l) + V(k,l)B_{i}(k,l) + N_{i}, i = L, R$$
(1)

where $H_i(k, l)$ and $B_i(k, l)$ are the acoustic transfer function from the target speech and interference source to the hearing aid user, respectively. S(k, l) and V(k, l) represent the speech and interference sources. N_i is the background diffuse noise propagating in all directions with equal power and random phase [3][4]. k and l denote the frequency bin and frame indices, respectively. The interference includes several directional talkers and additional transient noise sources, where a dominant noise source originates from a specific direction. Thus, the interference shows a high correlation between the left and right channel signals. On the other hand, background noise sources are often uncorrelated with each other except at low frequencies [3][4]. For the algorithm development, we assume that $E\{S(k,l)V^*(k,l)\} = 0$, $E\{S(k,l)N_i^*(k,l)\} = 0$ and $E\{N_i(k,l)N_j^*(k,l)\} = 0$ if $i \neq j$. Then, the auto-PSD of the binaural input signal is represented as

$$\Phi_X^i = E\{|X_i|^2\} = |H_i|^2 \Phi_S + |B_i|^2 \Phi_V + \Phi_N$$

= $\Phi_S^i + \Phi_V^i + \Phi_N, i = L, R.$ (2)

The frequency and frame indices, k and l, are omitted to simplify the notation for the rest of this paper. Because of the diffuse noise characteristics, the left and right diffuse noise sources have approximately equal PSDs [4], $\Phi_N \approx \Phi_N^L \approx \Phi_N^R$.

2.1. Bias due to the diffuse noise

Under the assumption that the direction of the target signal is exactly known, the noise-only reference signal for the left channel input can be obtained as

$$U_L = X_L - A_R X_R = \alpha_L V + N_L - A_R N_R \qquad (3)$$

where $A_R = H_L/H_R$ is the TFR of the target signal, and $\alpha_L = B_L - A_R B_R$. Using U_L , the target and interference signals contained in X_L can be separated by solving a minimum mean square error (MMSE) problem: $\min_{W_L} E\{|X_L - W_L^*U_L|^2\}$. The solution to this problem is obtained as

$$W_L^{o*} = \Phi_{UX}^L / \Phi_U^L, \tag{4}$$

$$\Phi_{UX}^L = E\{X_L U_L^*\} = B_L \alpha_L^* \Phi_V + \Phi_N \tag{5}$$

$$\Phi_U^L = E\{|U_L|^2\} = |\alpha_L|^2 \Phi_V + (1+|A_R|^2)\Phi_N.$$
(6)

Defining the error signal as $E_L = X_L - W_L^* U_L$, the target signal PSD can be estimated by computing the MMSE with optimum coefficients:

$$\Phi_E^L = E\{|E_L|^2\}_{W_L = W_L^o} = \Phi_X^L - |W_L^o|^2 \Phi_U^L$$
$$= \frac{\Phi_X^L \Phi_U^L - |\Phi_{UX}^L|^2}{\Phi_U^L} \approx \Phi_S^L.$$
(7)

The above equation can be found in [2], and a similar equation can be obtained from the interference estimation process in [1]. However, the estimate of target PSD given in Eq. (7) is biased because $\Phi_N \neq 0$. By applying Eqs. (2)(4)(6) to (7) and after some algebra, we show that the MMSE in Eq. (7) can be rewritten as

$$\Phi_E^L = \Phi_S^L + g \frac{\zeta + d}{\zeta + c} \Phi_N, \tag{8}$$

where $d = |A_R|^2/(|\alpha_L + B_L|^2 + |B_L|^2|A_R|^2)$, $c = (1 + |A_R|^2)/|\alpha_L|^2$, $g = (|\alpha_L + B_L|^2 + |B_L|^2|A_R|^2)/|\alpha_L|^2$, and

 $\zeta = \Phi_V / \Phi_N$ is the interference to noise ratio (INR). Thus the target PSD is always over-estimated, i.e., $\Phi_E^L > \Phi_S^L$, unless $\Phi_N = 0$. Fig. 1 shows the over-estimation bias in Eq. (8) due to the diffuse background noise power. The bias according to the SNR was measured as $\epsilon_{SNR} = \Phi_E^L / \Phi_S^L - 1 = \left(g\frac{\zeta+d}{\zeta+c}\right)(1/\eta)$, where $\eta = \Phi_S^L / \Phi_N$ denotes the signal-to-noise ratio (SNR), and the INR was set to $\zeta = -10$ dB. Additionally, the bias according to the INR was measured as $\epsilon_{INR} = (|W_L^o|^2 \Phi_U^L) / \Phi_V^L - 1 = \left(g'\frac{\zeta+d'}{\zeta+c'}\right)$, where $g' = (2Re(B_L\alpha_L^*) - |B_L|^2(1 + |A_R|^2))/(|B_L|^2|\alpha_L|^2)$, $d' = 1/(2Re(B_L\alpha_L^*) - |B_L|^2(1 + |A_R|^2))$, and $c' = |B_L|^2(1 + |A_R|^2)/(|B_L|^2|\alpha_L|^2)$. Thus, it can be observed that the bias is inversely proportional to both the SNR and INR. The bias can be significant, especially when the INR or SNR is low.



Fig. 1. Bias in the target PSD estimate in (7) according to (a) SNR $(\zeta = -10 \text{ dB})$ and (b) INR $(\eta = 0 \text{ dB})$.

2.2. Unbiased target PSD estimator

To alleviate the over-estimation bias problem, the noise reference signal should be obtained at a high signal-tointerference-plus-noise ratio (SINR) condition, which can be achieved by applying a conventional single-channel noise suppression algorithm [5][6] to the noise reference input U_L . If the background diffuse noise is successfully suppressed such that the diffuse noise power becomes insignificant, then the noise reference input can be written as

$$\hat{U}_L = \Xi\{U_L\} \approx \alpha_L V,\tag{9}$$

where $\Xi\{\cdot\}$ denotes the noise suppression process. Now, it is straightforward to see that the auto- and cross-PSDs of the diffuse noise-suppressed reference input \hat{U}_L are given by $\Phi_{\hat{U}}^L = |\alpha_L|^2 \Phi_V$ and $\Phi_{\hat{U}X}^L = B_L \alpha_L^* \Phi_V$. Thus, the solution to the MMSE problem of $\min_{W_L} E\{|X_L - \hat{W}_L^* \hat{U}_L|^2\}$ can be found as

$$\hat{W}_L^{o*} \approx \frac{B_L}{\alpha_L}.\tag{10}$$

Now, the optimal solution is no longer affected by diffuse noise. Then, the MMSE is obtained as

$$\hat{\Phi}_E^L = E\{|\hat{E}_L|^2\}|_{W_L = \hat{W}_L^o} = \Phi_X^L - \frac{|\Phi_{\hat{U}X}^L|^2}{\Phi_{\hat{U}}^L}.$$
 (11)

Because we can readily prove that $\hat{\Phi}_E^L = |H_L|^2 \Phi_S + \Phi_N$, the unbiased target PSD can be obtained from the MMSE of the prediction filter, as given by $\Phi_S^L = \hat{\Phi}_E^L - \Phi_N$ or

$$\hat{\Phi}_{S}^{L} = \left(\frac{\Phi_{X}^{L}\Phi_{\hat{U}}^{L} - |\Phi_{\hat{U}X}^{L}|^{2}}{\Phi_{\hat{U}}^{L}}\right) - \Phi_{N}.$$
 (12)

However, this unbiased solution requires a process of suppressing the diffuse noise in the noise reference input. Observing the relationship between the input and output of the diffuse noise suppression process $\Xi{\cdot}$, we can rewrite the auto- and cross-PSDs as

$$\Phi_{\hat{U}X}^L = \Phi_{UX}^L - \Phi_N \tag{13}$$

$$\Phi_{\hat{\mu}}^{L} = \Phi_{U}^{L} - (1 + |A_{R}|^{2})\Phi_{N}.$$
(14)

Thus, if Φ_N and A_R are known, the auto- and cross-PSDs of the diffuse noise-suppressed reference input are computed using Eqs. (13) and (14), and the unbiased target PSD is finally obtained using Eq. (12). The estimated target PSD in Eq. (12) is unbiased even under the presence of strong diffuse noise. This estimator can be further extended to achieve more accurate results at low frequencies by considering the coherence function of diffuse noise [2].

3. BINAURAL NOISE SUPPRESSION SYSTEM

Based on the proposed target PSD estimator, we design a binaural noise reduction system for hearing aids, which is shown in Fig.2. When noisy input signals enter the system, the target-free noise reference signals U_i are generated using the TFR. Then, a target PSD is calculated using the unbiased signal prediction model, in which the effect of diffuse noise is removed. Finally, spectral gains for the noise reduction are computed using the estimated target PSD. For a practical implementation, however, the TFR of the target signal and the diffuse noise PSD are required to be identified on-line.



Fig. 2. Block diagram of proposed binaural noise reduction system

3.1. Target TFR Estimation

For the target TFR estimation, we utilize the method in [7]. First, a coefficient vector maximizing the output SNR of the binaural hearing aids is obtained by computing the principal eigenvector from a generalized eigenvalue problem:

 $\mathbf{R}_X \mathbf{W}_{BF} = \lambda \mathbf{R}_N \mathbf{W}_{BF}$, where $\mathbf{R}_X = [\Phi_X^L \Phi_X^{LR}; \Phi_X^{LR}; \Phi_X^{R}]$, and $\mathbf{R}_N = [\Phi_N \ 0; \ 0 \ \Phi_N]$. If we assume that only the target signal and diffuse noise are present, the principal eigenvector is given by $\mathbf{W}_{BF} = \gamma \mathbf{R}_N^{-1} \mathbf{H} = (\gamma/\Phi_N) [H_L H_R]$ [7] where γ is an arbitrary complex constant and $\mathbf{H} = [H_L H_R]^T$ is the target transfer function vector. Therefore, the TFR of the target signal can be obtained as

$$\mathbf{W}_{TFR} = \frac{\mathbf{W}_{BF}}{W_{BF}(2,1)} = \begin{bmatrix} \frac{H_L}{H_R} & 1 \end{bmatrix}^T.$$
 (15)

In practice, the target TFR can be recursively estimated using the noisy input and diffuse noise PSDs, which are also recursively estimated. Any imperfect estimation of those PSDs causes an error in the estimated principal eigenvector [7] and in turn results in a distortion of the enhanced output. Because the accuracy of the eigenvector estimation is dependent on the input SNR [8], we use the enhanced output signal of the noise suppression system to find the principal eigenvector. Thus, the generalized eigenvetor problem is rewritten as $\mathbf{R}_{\hat{S}} \mathbf{W}_{BF} = \lambda \mathbf{R}_N \mathbf{W}_{BF}$, where $\mathbf{R}_{\hat{S}}$ is the correlation matrix of the enhanced output \hat{S} .

3.2. Diffuse noise PSD estimation

There have been a variety of algorithms for estimating the PSD of the background diffuse noise on-line. Recently, sophisticated algorithms based on signal prediction [9] and eigen-analysis [10] were proposed. These recent algorithms are capable of tracking fast time-varying noise spectra without latency. In this paper, we use the eigen-analysis-based noise PSD estimation algorithm [10] whose performance is known to be independent of both the direction of that target speech source as well as the SNR condition.

3.3. Spectral gain computation

The proposed system obtains an unbiased estimate of the target PSD $\hat{\Phi}_{S}^{i}$ by eliminating the effect of the diffuse noise PSD on the prediction filter. The so-obtained target PSD is used to compute real-valued spectral gains for the noise suppression, calculated as $G = \max(\sqrt{G_L \cdot G_R}, \sigma_{min})$, where $G_i = \min(\sqrt{\hat{\Phi}_{S}^{i}/\Phi_X^{i}}, 1), i = L, R$. The spectral gain floor used in this paper was $\sigma_{min} = 0.1$. The enhanced output signals are obtained by applying this real-valued spectral gain to both the left and right channels in a way such that $\hat{S}_i(k, l) =$ $X_i(k, l) \cdot G(k, l), i = L, R$. As a result, the noise components are suppressed while the binaural cues contained in the original inputs are preserved [1][2].

4. SIMULATION RESULTS

For the simulations, speech sentences were extracted from TIMIT databases and binaurally convolved with HRIR pairs



Fig. 3. Estimated target PSD at 0 dB SINR. Proposed algorithm at (a) wide band. (b) low frequency band (0-2000Hz). BIENR at (c) wide band. (d) low frequency band (0-2000Hz). Left channel noisy input PSD (gray solid line), clean speech PSD (black solid line), and estimated speech PSD (red dashed line).



Fig. 4. ImsegSNR in scenario-1. White noise with SINR (a) -5 dB, (b) 0 dB, (c) 5 dB, cafeteria noise with SINR (d) -5 dB, (e) 0 dB, (f) 5 dB

corresponding to the target and interference directions, respectively. In addition, a popping the champagne noise was applied as an additional directional transient noise. The white noise and recorded cafeteria noise were used as background diffuse noise. The noisy input signal was segmented into subframes of 1024 samples with 50% overlap using a sine window at a sampling rate of 22.05 kHz. In the first test scenario, a female target voice was placed in front of a hearing aid user (0°) , and two male interference voices were located at 270° and 60° , respectively, together with the champagne noise at 210°. Then, background diffuse noise was added. In the second test scenario, the target speech was moved to 270° to the left, and interferences were placed at 0° and 60° , respectively, with the same champagne noise direction. The performance of the proposed algorithm was compared with the previous techniques in [1] (TS-BASE/WF) and [2] (BIENR). BIENR was implemented with a time-domain Wiener prediction filter including a 60-sample causality delay.

We first tested the accuracy of the estimated target PSD using the proposed algorithm under the first test scenario. The result is shown in Fig.3 and compared with the result obtained using BIENR. The overall input SINR was 0 dB. The results clearly show that the proposed algorithm achieves very accurate target PSD estimation in the tested noise environment. Next, the performance of the the noise suppression system in



Fig. 5. ImsegSNR in scenario-2. White noise with SINR (a) -5 dB, (b) 0 dB, (c) 5 dB, cafeteria noise with SINR (d) -5 dB, (e) 0 dB, (f) 5 dB

Fig. 2 was assessed in both test scenarios. The segmental SNR improvement (ImsegSNR) was computed as the difference between a noisy input and an enhanced output signal according to the difference SINR and INR. The results are shown in Figs.4 and 5. The results show that the proposed noise suppression system attains better performance than the conventional algorithms and is significantly better than TS-BASE/WF method under low-SINR conditions. It should be noted that BIENR used diffuse noise suppression gain, so that it achieved high segmental SNR even at low SINR.

5. RELATION TO PRIOR WORK

The proposed algorithm has structural similarity to the previous work in [1]. However, the algorithm in [1] more concentrated on reducing multiple interferences. Therefore, bias is inevitable when diffuse noise is highly present. Additionally, the proposed algorithm has a similarity to the algorithm in [2] with regard to the use of a prediction filter. However, the signal prediction in [2] was also conducted without any consideration of the diffuse noise. In fact, it can be proven that both the signal prediction in [2] and the interference compensation in [1] are equivalent in obtaining the optimum coefficients. Thus, when the diffuse noise has relatively strong power, the target PSD obtained using the method in [1] and [2] are biased. In addition, the algorithm in [2] requires an explicit computation of the prediction error, but the proposed algorithm uses the prediction model only to derive the unbiased target PSD.

6. CONCLUSION

We proposed an unbiased target PSD estimator for hearing aids operating in an environment with complex noise. By obtaining the MMSE under a diffuse noise-free condition, the target PSD is obtained, unaffected by the diffuse noise. Experiments in binuaral noise suppression were conducted, and the results demonstrated the superior performance of the proposed algorithm over the previous technique, especially in low-SNR conditions.

7. REFERENCES

- [1] J. Li, S. Sakamoto, S. Hongo, M. Akagi, and Y. Suzuki, "Two-stage binaural speech enhancement with wiener filter for high-quality speech communication," *Speech Communication*, vol. 53, no. 5, pp. 677–689, April 2011.
- [2] A.H. Kamkar-Parsi and M. Bouchard, "Instantaneous binaural target PSD estimation for hearing aid noise reduction in complex acoustic environments," *IEEE Trans. on Instrum. and Meas.*, vol. 60, no. 4, pp. 1141– 1154, April 2011.
- [3] I.A. McCowan and H. Bourlard, "Microphone array post-filter based on noise field coherence," *IEEE Trans.* on Speech and Audio Processing, vol. 11, no. 6, pp. 709 – 716, Nov. 2003.
- [4] H.R. Abutalebi, H. Sheikhzadeh, R.L. Brennan, and G.H. Freeman, "A hybrid subband adaptive system for speech enhancement in diffuse noise fields," *IEEE Signal Process. Lett.*, vol. 11, no. 1, pp. 44 – 47, Jan. 2004.
- [5] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. on Speech and Audio Processing*, vol. 9, no. 5, pp. 504 –512, July 2001.
- [6] I. Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," *IEEE Trans. on Speech and Audio Processing*, vol. 11, no. 5, pp. 466–475, Sept. 2003.
- [7] A. Krueger, E. Warsitz, and R. Haeb-Umbach, "Speech enhancement with a GSC-like structure employing eigenvector-based transfer function ratios estimation," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 19, no. 1, pp. 206–219, Jan. 2011.
- [8] Y. Hou and Y. Shen, "An improved signal subspace decomposition method in low SNR case," *International Symposium on Systems and Control in Aeronautics and Astronautics (ISSCAA), Harbin*, pp. 143–146, June 2010.
- [9] A.H. Kamkar-Parsi and M. Bouchard, "Improved noise power spectrum density estimation for binaural hearing aids operating in a diffuse noise field environment," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 17, no. 4, pp. 521–533, May 2009.
- [10] Y. Ji, Y.C Park, D.W Kim, and J. Sohn, "Robust noise PSD estimation for binaural hearing aids in time-varying diffuse noise field," *IEEE Intern. Conf. on Acoust., Speech, Signal Process. (ICASSP), Vancouver, Canada*, pp. 7264 – 7268, May 2013.