ROBUST NOISE PSD ESTIMATION FOR BINAURAL HEARING AIDS IN TIME-VARYING DIFFUSE NOISE FIELD

Youna Ji, Young-cheol Park

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

Dong-wook Kim, Junil Sohn

Samsung Advanced Institute of Technology
Ambient Platform Group
Suwon, Korea

ABSTRACT

In this paper, we present an unsupervised noise PSD estimation algorithm for binaural hearing aids in a time-varying diffuse noise field. It is shown that the noise PSD can be obtained from the eigenvalues of the input covariance matrix together with the noise coherence function effective at low frequencies. To reduce the estimation bias due to fast smoothing, pre- and post-compensation methods are proposed. The proposed algorithm is able to track non-stationary noise PSD without tracking delay or underestimation problems. Its performance is independent of the target speech direction and input SNR. Results of the objective parameter evaluation demonstrate the superiority of the proposed algorithm over conventional techniques.

Index Terms— binaural noise estimation, binaural hearing aids, recursive averaging, noise coherence

1. INTRODUCTION

Noise reduction algorithms in hearing aids are crucial for improving speech intelligibility and quality under background noise. Since accuracy of the noise power spectral density (PSD) directly affects the performance of a speech enhancement system, robust estimation of noise PSD from available noisy signals has been an important issue.

The minimum statistics (MS) technique [1] was established based on the fact that the minimum power of noisy speech is approximately the same as the power level of the noise. However, this technique inherently has some tracking latency because of the searching window [2]. In a recent study, an effective noise PSD estimation algorithm based on a prediction model [2] was suggested. It was shown that this algorithm can track the true noise PSD without tracking latency and underestimation problems at low frequencies. However, there are still problems in cases in which the target speech source is located in a non-frontal direction. Although the problems can be solved by inserting a causality delay between channel signals, the delay often causes overestimation of noise PSD, which is mainly due to the loss of accuracy in computing the prediction errors.

Regarding the noise PSD estimation, there are some implementation issues that are encountered in fast time-varying noise environments. First of all, the characteristics of environmental noises can be time-varying and dependent on the position of surrounding sources [3]. In such an environment, use of the analytic coherence model [2][4][5] is problematic. Another practical issue is raised by smoothing with a short time constant. Auto- and cross-PSD can be efficiently estimated using 1st-order recursive averaging. In rapidly changing noise environments, fast smoothing is necessary to obtain the noise PSD without latency, but it can result in bias with PSD estimates [6].

In this paper, we propose a new unsupervised noise PSD estimator for binaural hearing aids operating in a fast time-varying diffuse noise field. It will be shown that the noise PSD can be estimated directly from the eigenvalues of the input covariance matrix, in cooperation with the noise coherence function. To obtain the noise coherence without *a priori* knowledge about the noise field, an on-line estimation technique will be employed. Also, to alleviate the spectral error due to fast smoothing, the cross-PSD in the covariance matrix will be compensated for prior to the eigen-analysis. Later, the spectral error is further reduced by compensating for the eigenvalues in noise-only regions.

2. DIFFUSE NOISE PSD ESTIMATION ALGORITHM

The left- and right-channel noisy input signals are represented in the frequency-domain as

$$X_i(k,l) = S(k,l)H_i(k,l) + N_i(k,l), i = L, R, \quad (1)$$

where $H_i(k,l)$ is the acoustic path from the target speech source to the hearing aid user, $S(k,l)$ is the speech source and $N_i(k,l)$ is the diffuse noise signal. k and l denote the frequency bin and frame indices, respectively. The $N_i(k,l)$ are considered to propagate diffuse noise in all directions simultaneously with equal power and random phase [7][4]. For the algorithm development, we assume that $E\{S(k,l)N_i^*(k,l)\} = 0$ and the left and right diffuse noises have approximately equal PSD [2][7].

2.1. Unsupervised Diffuse Noise PSD Estimation

The covariance matrix of noisy inputs in (1) is obtained as $\mathbf{R} = \begin{pmatrix} \Phi_X^{LL} & \Phi_X^{LR} \\ \Phi_X^{RL} & \Phi_X^{RR} \end{pmatrix}$ where $\Phi_X^{ij} = E\{X_i X_j^*\}, i, j = L/R$, or equivalently expressed as

$$\mathbf{R} = \begin{pmatrix} |H_L|^2\Phi_S + \Phi_N & H_L H_R^*\Phi_S + \Phi_N^{LR} \\ H_R H_L^*\Phi_S + \Phi_N^{RL} & |H_R|^2\Phi_S + \Phi_N \end{pmatrix} \quad (2)$$

where $\Phi_S = E\{|S|^2\}$ and $\Phi_N = E\{|N_L|^2\} = E\{|N_R|^2\}$ are target speech source and noise auto PSDs respectively, and $\Phi_N^{ij} = E\{N_i N_j^*\}$ is the noise cross PSD. Here, we omitted the frame and frequency bin indices for convenience. Noises measured in a diffuse field have significant correlation, especially at low frequencies [8]. To account for the correlation of noise fields, an analytic coherence model is often used, which is given by $\Gamma_N(k) = \text{sinc}(2\pi kd/c)$ [2][4] where c is the speed of sound, and d is distance between the measurement points. Using the coherence model, the noise cross PSD can be expressed as $\Phi_N^{LR} \approx \Gamma_N \Phi_N$.

The eigenvalues of the covariance matrix in (2) can be computed by solving the characteristic equation. The two eigenvalues are then given by

$$\lambda_{1,2} = \frac{((|H_L|^2 + |H_R|^2)\Phi_S + 2\Phi_N) \pm \sqrt{\Delta}}{2}, \quad (3)$$

where $\Delta = (|H_L|^2 + |H_R|^2)^2\Phi_S^2 + 2\Gamma_N\Phi_N\Phi_S(2H_L H_R^* + 2H_L^* H_R) + 4\Gamma_N^2\Phi_N^2$. It is known that the low frequency interaural level differences (ILDs) are negligible [9], i.e., $|H_L| \approx |H_R|$, and ignoring the phase difference between the acoustic paths, we yield $2H_L H_R^* \approx |H_L|^2 + |H_R|^2$. Applying this approximation, we obtain $\Delta \approx ((|H_L|^2 + |H_R|^2)\Phi_S + 2\Gamma_N\Phi_N)^2$. Thus, the second (smaller) eigenvalue can be expressed as $\lambda_2 \approx (1 - \Gamma_N)\Phi_N$, from which the noise PSD is obtained: $\hat{\Phi}_N \approx \lambda_2/(1 - \Gamma_N)$. The above procedure is effective only at low frequencies, typically lower than 500Hz, where ILDs are negligible. At high frequencies the coherence of the diffuse noise is negligibly small, i.e., $|\Gamma_N| \approx 0$. Therefore, the noise PSD can be estimated as $\Phi_N \approx \lambda_2$ at high frequencies. In summary, the noise PSD is calculated as

$$\hat{\Phi}_N(k,l) \approx \begin{cases} \frac{\lambda_2(k,l)}{1-\Gamma_N(k)} & \text{if } k < k_c \\ \lambda_2(k,l) & \text{if } k \geq k_c, \end{cases} \quad (4)$$

where k_c denotes the bin index corresponding to the highest frequency at which the approximation is effective.

2.2. On-line Estimation of Noise Coherence

The coherence between two microphones changes, unlike the analytic model $\Gamma_N(k)$, in which an object is in the line-of-sight [5]. Moreover, the noise coherence in a practical situation is slowly time varying according to the position of the sound sources and the acoustic environment [3]. Thus, it is beneficial to update the noise coherence on-line in accordance with the variation of the acoustic environment. Given two hypotheses, H_0 and H_1, which represent speech absent and speech present, respectively, the noise coherence can be updated only during the absence of speech, as given by

$$\tilde{\Gamma}_N(k,l) = \begin{cases} \tilde{\Gamma}_N(k,l-1) & : H_1(k,l) \\ \beta\tilde{\Gamma}_N(k,l-1) + (1-\beta)|\Gamma_X(k,l)| & : H_0(k,l), \end{cases} \quad (5)$$

where β is a smoothing parameter and Γ_X represents inter-channel coherence given by $\Gamma_X = \Phi_X^{LR} \Big/ \sqrt{\Phi_X^{LL}\Phi_X^{RR}}$.

Previously, energy ratio-based techniques have been widely used to determine the presence of speech [10]. In [10], the speech absent probability was determined based on a combination of local, global and frame *a priori* SNRs. In this paper, the energy-ratio is measured as a form of eigen-ratios and it is used to determine the presence of speech. To this end, the ratio between total powers of first (larger) and second (smaller) eigenvalues within the band b is calculated:

$$\varphi(b,l) = \frac{\sum_{k\in b}\lambda_1(k,l)}{\sum_{k\in b}\lambda_2(k,l)}. \quad (6)$$

Then, the speech present region is determined as by comparing $\varphi(b,l)$ with a threshold δ, as given by

$$\begin{aligned} \varphi(b,l) < \delta \quad &: \text{speech absent } (H_0) \\ \varphi(b,l) \geq \delta \quad &: \text{speech present } (H_1). \end{aligned}$$

Assuming that left and right channel diffuse noises are uncorrelated. We can show that the eigen-ratio is an alternative measure of *a priori* SNR $\xi(k,l) = \frac{\Phi_S(k,l)}{\Phi_N(k,l)}$:

$$\frac{\lambda_1(k,l)}{\lambda_2(k,l)} \approx \xi(k,l)(|H_L(k,l)|^2 + |H_R(k,l)|^2) + 1. \quad (7)$$

Thus, $\varphi(b,l)$ is an appropriate metric for measuring energy-ratios that can be used to determine the presence of speech. An example of noise coherence tracking in the 150-400 Hz

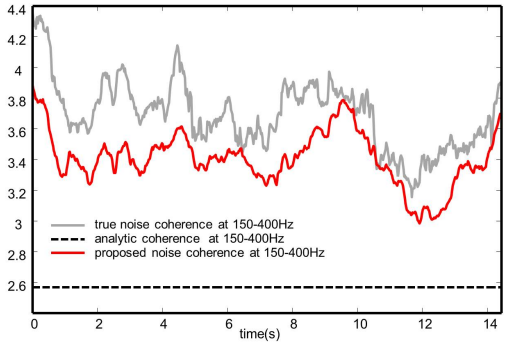

Fig. 1. On-line update of coherence: band-coherence obtained from true noise (grey), modified coherence (red) and analytic model (dashed).

frequency region was obtained using the proposed on-line estimation shown in Fig.1. We used $\beta = 0.95$ and cafeteria

signals as background noise . It can be seen that the on-line estimation method tracks the true noise coherence comparing analytic model.

3. COMPENSATION FOR UNDERESTIMATION OF NOISE PSD

3.1. Compensation of cross-power spectral density measured by fast smoothing

In practice, auto- and cross-PSD of the input noisy signals are often obtained using a 1st-order recursive approximation:

$$\tilde{\Phi}_X^{ij}(k,l) = \alpha\tilde{\Phi}_X^{ij}(k,l-1) + (1-\alpha)X_i(k,l)X_j^*(k,l) \quad (8)$$

where $\alpha \in [0,1]$ is the smoothing factor which controls trade off relationships between the fast capturing of time-varying statistics of the signals and obtaining a robust low-variance estimate of the spectrum [6][11]. The cross-PSD tends to be overestimated in a fast smoothing environment, and in turn it can result in serious underestimation of the final noise PSDs. Nevertheless, fast smoothing is commonly required to track the spectrum of non-stationary noises. Estimation errors also occur for the computed auto-PSD but these errors are typically small compared to those seen in the cross-PSD [11]. In [11], a method of compensating cross-correlation under fast smoothing conditions was suggested by linearly expanding inter-channel correlation coefficients. This approach is effective in speech absent regions but it also reduces high coherence regions in which target speech is likely to be present, which can cause serious speech distortions.

In this paper, we use a new pre-compensation method for cross-PSD. We first express the cross-PSD between the input noisy signals using the coherence as given by

$$\tilde{\Phi}_X^{ij}(k,l) = \hat{\Gamma}_X(k,l)\sqrt{\Phi_X^{ii}(k,l)\Phi_X^{jj}(k,l)}, i,j = L/R. \quad (9)$$

Next, through intensive tests and statistical analysis, we experimentally derived a compensation function for a magnitude square coherence(MS-coherence) function. The resulting compensation rule is given by

$$|\hat{\Gamma}_X(k,l)|^2 = \psi(k,l)|\Gamma_X(k,l)|^2, \quad (10)$$

where

$$\psi(k,l) = \frac{1}{1+exp(a(1-|\Gamma_X(k,l)|^2)-b\alpha)}. \quad (11)$$

The scaling factor ψ is formed as a sigmoid function that has nonlinear and bounded characteristics. In this paper, $a = 14$ and $b = 12$ were selected to provide the best compensation. The sigmoid function obviously separates the high and low correlation cases. It has a clear effect of reducing the coherence overestimation, which in turn results in a reduction of cross-PSD errors through Eq. (9). The compensated coherence $\hat{\Gamma}_X(k,l)$ can also be applied to the on-line update of the noise coherence shown in (5).

3.2. Post-compensation for Under-estimation of Noise PSD

When fast smoothing is used, the estimated diffuse noises can often have unequal power. Thus, eigenvalues obtained in speech pause periods cannot produce exact noise PSD. To overcome this problem, we propose a post-compensation method for recovering the true noise PSD by combining the two eigenvalues, as given by the following equation:

$$\lambda_c(b,l) = \alpha_n\lambda_2(b,l) + (1-\alpha_n)\lambda_1(b,l), \quad (12)$$

where $b \in [0, 2\pi]$ for H_0 and $b \in [0, \omega_c]$ for H_1 The frequency index ω_c is a cutoff corresponding to an ultra low-frequency band in which the speech signal is not present. Typically $\omega_c \approx 100Hz$. Fig. 2 shows distributions of left and right channel true noise PSDs together with eigenvalues. The noisy input signal was 5dB SNR and a smoothing factor of $\alpha = 0.65$ was used. PSDs were accumulated from $1500Hz$ to $4000Hz$ band. Fig. 2 shows that the second eigenvalue, determined as the noise PSD, in a speech absent region is significantly smaller than the true noise PSD, but the compensation using (12) restores the energy distribution of noise PSD estimates close to that of true noise PSDs. Now, the compensated noise PSD can be obtained using (4) by replacing $\lambda_2(k,l)$ with $\lambda_c(k,l)$.

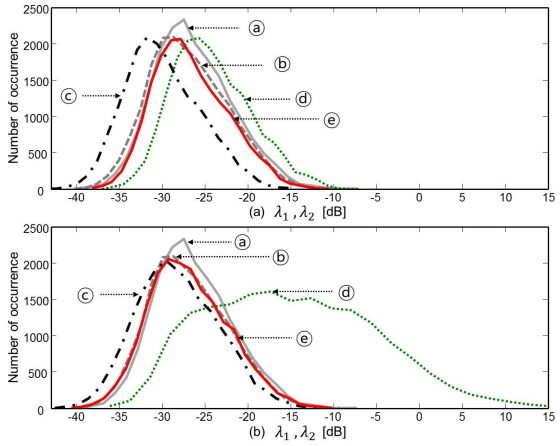

Fig. 2. Distributions (a) in speech absent region and (b) speech present region (ⓐ left-channel and ⓑ right-channel true noise PSDs, ⓒ 1st and ⓓ 2nd eigenvalues. ⓔ noise PSD estimate obtained using (12)).

4. SIMULATIONS

Simulations for assessing the proposed noise PSD estimation algorithm were performed. Selected speech sentences from TIMIT databases were binaurally convolved with HRIR pairs corresponding to target directions, and a recoded cafeteria diffuse noise was added to the binaural speech sentences at various SNRs. The sampling frequency was $16kHz$. The FFT frame length was $32ms$ with a 50% overlap and a Sine

Table 1. Objective parameters

	segSNR			LSD	
Angle	Noisy	ImNPSD	Proposed	mNPSD	Proposed
0°	-0.59	3.45	4.24	3.42	2.79
	3.43	6.12	7.25	4.15	2.95
	7.72	9.19	10.59	5.33	3.68
270°	-0.40	1.64	2.71	2.66	2.03
	3.64	3.78	5.31	3.13	1.98
	7.73	6.22	8.34	4.03	2.10

window was applied to each frame before FFT. The performance of the proposed algorithm was compared with the previous technique in [2], referred to as improved noise PSD (ImNPSD). ImNPSD was implemented with a time-domain Wiener prediction filter including 40 sample causality delays.

Snapshots of the estimated noise PSDs are shown in Fig. 3. The results were obtained with a target speech source located at 270° to the left of the listener and the SNR was 5dB. ImNPSD shows spectral errors, which is mainly due to the causality delay used for the Wiener prediction filter, which was not optimal. In large interaural time difference (ITD) situations, ImNPSD produced significant estimation errors especially for unvoiced regions. In such situations, the signal power of one-channel is usually smaller than the other channel because of the head-shading effect; therefore, the prediction model becomes inaccurate. On the other hand, the proposed method produces robust noise PSD estimates regardless of target speech direction and does not require delays. Using the estimated noise PSDs, a noise reduction based on the Wiener filter was conducted. The decision-directed approach was used to estimate *a priori* SNR. After noise reduction, objective parameters such as segmental SNR (segSNR) and log spectral distortion (LSD) of the target speech were measured. The segSNR and LSD assessments are summarized in Table 1. The results show that the proposed technique always attains higher segSNR and lower LSD than ImNPSD, which indicates that it obtains the true noise PSD with higher accuracy than the ImNPSD method.

Fig. 4 shows tracking performance of the proposed noise PSD estimation algorithm. The proposed method with compensation shows excellent tracking performance.

5. RELATION TO PRIOR WORK

The proposed noise PSD estimation algorithm can be compared with the recent work in [2] in which a noise PSD estimator based on the Wiener prediction filter was suggested. Major advantages of the proposed algorithm in comparison to [2] is that its performance is independent of both the direction of the target speech source as well as the SNR condition. Moreover, the noise coherence model is updated on-line, which is also advantageous when noise characteristics are time-varying. Pre-compensation of cross-PSD in this study

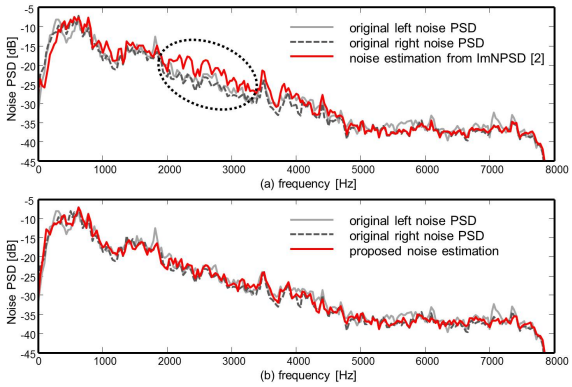

Fig. 3. Estimated noise PSDs when a target speech was 270° from frontal direction. SNR was 5dB. Gray solid line and dashed line represent left and right channel true noise PSDs, respectively. Red solid lines are the estimated noise PSD using (a) ImNPSD and (b) proposed method

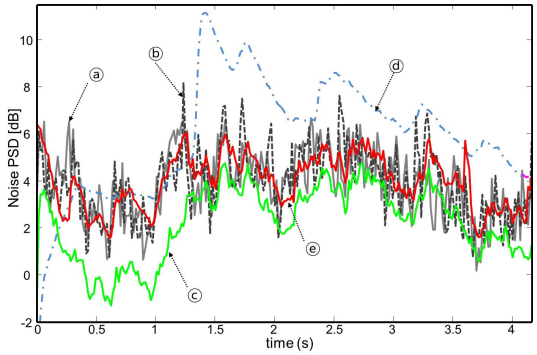

Fig. 4. Tracking performance according to smoothing actor: ⓐ left channel noise PSD, ⓑ right channel noise PSD, estimated noise PSDs with ⓒ $\beta = 0.65$ without scaling, ⓓ $\beta = 0.93$ without scaling, ⓔ $\beta = 0.65$ with compensation methods in (10) and (12).

was motivated by the previous work in [11]. However, the previous work cannot properly compensate for the high coherence region, while the proposed method can effectively reduce the spectral error of noise PSD without distorting the target speech signal. Finally, the post-compensation of eigenvalues in this study can be compared with the previous work in [3] in which the noise eigenvalues were modified to satisfy the rank-1 condition. The eigenvalue modification in this study is to compensate for underestimation of noise PSD in a fast smoothing condition.

6. CONCLUSION

In this paper, a new unsupervised noise PSD estimator based on the eigen-structure of the binaural noisy input covariance matrix was proposed. Via on-line estimation of noise coherence, and by compensating spectral parameters, the proposed algorithm is able to obtain accurate noise PSDs even under fast smoothing conditions. Simulation results demonstrated superior performance of the proposed algorithm over the previous techniques.

7. REFERENCES

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

[2] A.H. Kamkar-Parsi and M. Bouchard, "Improved noise power spectrum density estimation for binaural hearing aids operating in a diffuse noise field environment," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 17, no. 4, pp. 521 –533, may 2009.

[3] Gibak Kim and Nam Ik Cho, "Frequency domain multi-channel noise reduction based on the spatial subspace decomposition and noise eigenvalue modification," *Speech Communication*, vol. 50, no. 5, pp. 382 – 391, 2008.

[4] I.A. McCowan and H. Bourlard, "Microphone array post-filter based on noise field coherence," *Speech and Audio Processing, IEEE Transactions on*, vol. 11, no. 6, pp. 709 – 716, nov. 2003.

[5] M. Jeub and P. Vary, "Binaural dereverberation based on a dual-channel wiener filter with optimized noise field coherence," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, march 2010, pp. 4710 –4713.

[6] B.N.M. Laska, M. Bolic and, and R.A. Goubran, "Coherence-assisted wiener filter binaural speech enhancement," *Instrumentation and Measurement Technology Conference (I2MTC), 2010 IEEE*, pp. 876 –881, may 2010.

[7] H.R. Abutalebi, H. Sheikhzadeh, R.L. Brennan, and G.H. Freeman, "A hybrid subband adaptive system for speech enhancement in diffuse noise fields," *Signal Processing Letters, IEEE*, vol. 11, no. 1, pp. 44 – 47, jan. 2004.

[8] Matthias Dorbecker and Stefan Ernst, "Combination of two–channel spectral subtraction and adaptive wiener post–filtering for noise reduction and dereverberation," in *in European Signal Processing Conference (EUSIPCO–96*, 1996.

[9] V. Ralph Algazi, Carlos Avendano, and Richard O. Duda, "Elevation localization and head-related transfer function analysis at low frequencies," *The Journal of the Acoustical Society of America*, vol. 109, no. 3, pp. 1110–1122, 2001.

[10] I. Cohen, "Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator," *Signal Processing Letters, IEEE*, vol. 9, no. 4, pp. 113 –116, april 2002.

[11] Juha. Merimaa, Michael M. Goodwin, and Jean-Marc Jot, "Correlation-based ambience extraction from stereo recordings," in *Audio Engineering Society Convention 123*, oct. 2007.