

## SOUND-LOCALIZATION-PRESERVED BINAURAL MMSE STSA ESTIMATOR WITH EXPLICIT AND IMPLICIT BINAURAL CUES

H. Saruwatari<sup>†</sup>, R. Wakisaka<sup>†</sup>, K. Shikano<sup>†</sup>, F. Mustiere<sup>‡</sup>, L. Thibault<sup>‡</sup>, H. Najaf-Zadeh<sup>‡</sup>, M. Bouchard<sup>\*†</sup>

<sup>†</sup>Nara Institute of Science and Technology, 630-0192 Nara, Japan,

<sup>‡</sup>Communications Research Centre, Ottawa, Canada, and <sup>\*</sup>University of Ottawa, Ottawa, Canada.

### ABSTRACT

In this paper, we address some variations of the source-localization-preserved MMSE STSA estimator used for binaural hearing aids. In our previous work, the sound-localization-preserved MMSE STSA estimator with ICA-based noise estimation has been proposed. However, this conventional method is based on an approximated optimization criterion and does not use binaural cues, resulting in poor noise reduction performance. To solve this problem, we propose two methods: a multichannel MMSE STSA estimator with explicit binaural cues, and a sound-localization-preserved generalized MMSE STSA estimator with different speech priors for the left and right channels as implicit binaural cues. From the results of objective and subjective evaluation, we confirm that the noise reduction performance is improved using the proposed method.

**Index Terms**— Hearing aids, ICA, MMSE STSA estimator, localization, generalized gamma distribution

### 1. INTRODUCTION

In recent years, the number of applications of speech communication systems has increased. These systems, however, always suffer from the problem of a deterioration of speech quality under adverse noise conditions. Therefore, many noise reduction methods have been actively studied. In this paper, we address a noise reduction technique for binaural hearing-aid systems [1, 2, 3] by evaluating the quality of a speech-enhanced signal according to the human perceptual impression as well as by the amount of noise reduction.

The basic theory of the minimum mean square error (MMSE) short-time spectral amplitude (STSA) estimator has been presented by Ephraim et al. [4] for the optimal identification of the target speech amplitude spectrum in the MMSE sense, ignoring phase spectrum information. The MMSE STSA estimator has a good balance between noise reduction ability and low speech distortion. In previous works, the sound-localization-preserved MMSE STSA estimator with dynamic noise estimation based on independent component analysis (ICA) [5] has been proposed by the authors [6] for achieving nonstationary noise reduction. In [6] as well as [3], it is shown that sound localization can be preserved using a common spectral gain for noise reduction at each channel of the left and right ears. However, the conventional method proposed in [6] was based on an approximated optimization criterion and did not use binaural cues to estimate the optimal

common spectral gain for each channel. This often causes deterioration in the noise reduction performance.

To improve the noise reduction performance of the conventional method, we propose two methods in this paper. First, we introduce a multichannel MMSE STSA estimator to provide a strict solution without any approximation. This method greatly improves the sound quality, but is not practical because an accurate head-related transfer function (HRTF) of each user is required as explicit binaural cues. Therefore, secondly, we propose a sound-localization-preserved *generalized* MMSE STSA estimator. In this method, we replace the original MMSE STSA estimator in the conventional method with a generalized MMSE STSA estimator, to which we apply different speech priors for the left and right channels as implicit binaural cues. From the results of objective and subjective evaluation, we confirm that the noise reduction performance is improved using the proposed method.

### 2. SIGNAL MIXTURE MODEL

We consider an acoustic mixing model with two microphones, i.e., two earphones, and assume that the observed signal contains only one target speech signal, which can be regarded as a point source, and an additive noise signal. This additive noise signal cannot be regarded as a point source. Hereafter, the observed signal vector in the time-frequency domain,  $\mathbf{x}(f, \tau) = [x_L(f, \tau), x_R(f, \tau)]^T$ , is given by

$$\mathbf{x}(f, \tau) = \mathbf{h}(f)s(f, \tau) + \mathbf{n}(f, \tau), \quad (1)$$

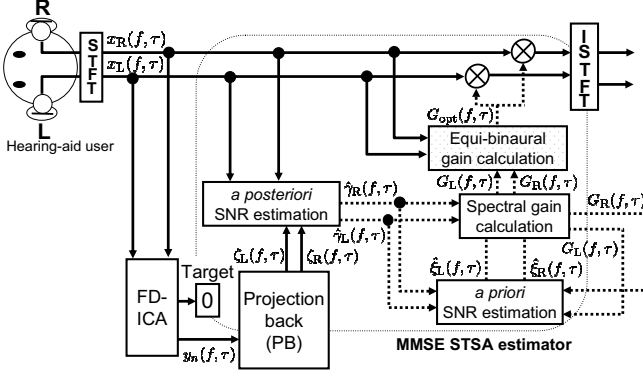
where  $f$  is the frequency bin number,  $\tau$  is the time-frame index,  $\mathbf{h}(f) = [h_L(f), h_R(f)]^T$  is the column vector of transfer functions between the target source and each earphone,  $s(f, \tau)$  is the target speech signal component, and  $\mathbf{n}(f, \tau) = [n_L(f, \tau), n_R(f, \tau)]^T$  is the column vector of the additive noise signal. Throughout this paper, the subscripts (or superscripts) L and R represent the signals obtained at the left and right ears, respectively.

### 3. CONVENTIONAL METHOD

#### 3.1. MMSE STSA Estimator with ICA-Based Noise Estimation

Figure 1 shows a block diagram of the conventional method consisting of an MMSE STSA estimator-based primary path and a reference path for ICA-based noise estimation [6]. The noise component estimated by ICA is used to determine the *a posteriori* signal-to-noise ratio (SNR) and the optimal spectral gain applied to the L and R channel signals, neglecting

This work was partly supported by JST Core Research of Evolutional Science and Technology (CREST), Japan.



**Fig. 1.** Conventional sound-localization-preserved MMSE STSA estimator with ICA-based noise estimation [6].

phase information; this enables the method to realize error-robust noise reduction.

First, in the reference path, we perform signal separation and noise estimation using ICA [5]. Due to the limitation of the paper space, we omit the detailed algorithm to estimate noise (see [7] for further details). Next, using the noise components estimated by ICA, we derive the gain function optimized on the basis of the MMSE STSA estimation criterion. Hereinafter, for convenience, the subscript  $* \in \{L, R\}$  is used to indicate the index of channels. The estimated a posteriori SNR  $\hat{\gamma}_*(f, \tau)$  is obtained by

$$\hat{\gamma}_*(f, \tau) = |x_*(f, \tau)|^2 \{E\{|\zeta_*(f, \tau)|^2\}_{\tau-\tau_{th}}^\tau}\}^{-1}, \quad (2)$$

where  $\zeta_*(f, \tau)$  is the noise component estimated by ICA,  $\tau_{th}$  is a smoothing parameter denoting a certain time frame window, and  $E\{\cdot\}_A^B$  denotes the expectation operator from  $A$  to  $B$ . Note that we can momentarily estimate the instantaneous a posteriori SNR (2) by utilizing the noise signal estimated by ICA, in contrast to the case of using the original MMSE STSA estimator [4]. Therefore, we consider that our proposed method can suppress nonstationary noise more efficiently than the conventional MMSE STSA estimator.

Next, using (2), the *a priori* SNR  $\hat{\xi}_*(f, \tau)$  is estimated as

$$\hat{\xi}_*(f, \tau) = \alpha \hat{\gamma}_*(f, \tau - 1) G_*^2(f, \tau - 1) + (1 - \alpha) P[\hat{\gamma}_*(f, \tau) - 1], \quad (3)$$

where  $\alpha$  is the weighting factor of the decision-directed estimation,  $G_*(f, \tau)$  is a spectral gain function, and  $P[\cdot]$  is a flooring function in which a negative input is floored to zero. Also, the spectral gain function is defined by [4]

$$G_*(f, \tau) = \Gamma(1.5) \frac{\sqrt{\nu_*(f, \tau)}}{\hat{\gamma}_*(f, \tau)} \exp(\nu_*(f, \tau)/2) \cdot [\{1 + \nu_*(f, \tau)\} I_0(\nu_*(f, \tau)/2) + \nu_*(f, \tau) I_1(\nu_*(f, \tau)/2)], \quad (4)$$

where  $\Gamma(\cdot)$  denotes the gamma function and  $I_0(\cdot)$  and  $I_1(\cdot)$  denote modified Bessel function of the zeroth and first order, respectively. Moreover,  $\nu_*(f, \tau)$  is defined by

$$\nu_*(f, \tau) = \hat{\xi}_*(f, \tau) \hat{\gamma}_*(f, \tau) \{1 + \hat{\xi}_*(f, \tau)\}^{-1}. \quad (5)$$

Finally, noise reduction is carried out as follows:

$$z_*(f, \tau) = G_*(f, \tau) x_*(f, \tau), \quad (6)$$

where  $z_*(f, \tau)$  is the final output of this method for both ears.

### 3.2. Estimation of Equi-Binaural Optimal Spectral Gain

The results in [6, 3] indicate that it is essential to apply *equivalent* spectral gains for the L and R channels to increase the localization accuracy. Therefore, to obtain the optimal spectral gain that maintains the localization accuracy, we introduce the spectral gain that minimizes the residual noise power in terms of the MMSE under the condition that the spectral gains are equivalent in both channels. Hereafter, we call this gain the *equi-binaural optimal spectral gain*.

The derivation of the equi-binaural optimal spectral gain can be formulated as the minimization problem of the following error  $e$ :

$$e = E \left[ \{|h_L(f)s(f, \tau) - G(f, \tau)x_L(f, \tau)|\}^2 + \{|h_R(f)s(f, \tau) - G(f, \tau)x_R(f, \tau)|\}^2 \right], \quad (7)$$

where  $G(f, \tau)$  is the equi-binaural spectral gain, which is considered as a variable. This problem can be *approximately* reformulated as [6]

$$G_{opt}(f, \tau) \simeq \underset{G(f, \tau)}{\operatorname{argmin}} E \left[ \{(G(f, \tau) - G_{L_{opt}}(f, \tau))x_L(f, \tau)\}^2 + \{(G(f, \tau) - G_{R_{opt}}(f, \tau))x_R(f, \tau)\}^2 \right], \quad (8)$$

where  $G_{L_{opt}}(f, \tau)$  and  $G_{R_{opt}}(f, \tau)$  are the L- and R-channel optimal spectral gains, respectively, given by (4). The solution of (8) is given by

$$G_{opt}(f, \tau) = \frac{G_{L_{opt}}(f, \tau)|x_L(f, \tau)|^2 + G_{R_{opt}}(f, \tau)|x_R(f, \tau)|^2}{|x_L(f, \tau)|^2 + |x_R(f, \tau)|^2}. \quad (9)$$

## 4. PROPOSED METHOD I: HRTE-INFORMED MULTICHANNEL MMSE STSA ESTIMATOR

### 4.1. Multichannel MMSE STSA Estimator

In the conventional method [6], the equi-binaural optimal spectral gain is approximately estimated using optimal spectral gains independently obtained from single-channel MMSE STSA estimation for each channel. Thus, the conventional method does not use binaural cues such as the HRTF. This fact often causes serious degradation of the noise reduction performance, particularly in the case that the target speech source is laterally located on the right- (left-) hand side. Therefore, in this section, we introduce the *oracle* method using the multichannel MMSE STSA estimator [8], which explicitly uses the HRTF in the estimation of the target speech instead of single-channel MMSE STSA estimators for each channel. This method can give the equi-binaural optimal spectral gain without approximation.

The multichannel MMSE STSA estimator can be viewed as a cascade of minimum variance distortionless response (MVDR) beamforming and the single-channel MMSE STSA estimator [8]. The output of MVDR beamforming is given by

$$Y(f, \tau) = \frac{\mathbf{h}(f)^H \Sigma_N(f, \tau)^{-1} \mathbf{x}(f, \tau)}{\mathbf{h}(f)^H \Sigma_N(f, \tau)^{-1} \mathbf{h}(f)}, \quad (10)$$

where  $\Sigma_N(f, \tau)$  is a noise covariance matrix momentarily estimated by ICA, which is defined as

$$\Sigma_N(f, \tau) = \mathbb{E}\{[\zeta_L(f, \tau), \zeta_R(f, \tau)]^H [\zeta_L(f, \tau), \zeta_R(f, \tau)]\}_{\tau-\tau_{th}}^{\tau} \quad (11)$$

Then, the a posteriori SNR for  $Y(f, \tau)$  is calculated as

$$\hat{\gamma}_Y(f, \tau) = |Y(f, \tau)|^2 \mathbf{h}(f)^H \Sigma_N(f, \tau)^{-1} \mathbf{h}(f). \quad (12)$$

Next, using (12), the a priori SNR estimate  $\hat{\xi}_Y(f, \tau)$  is given as

$$\hat{\xi}_Y(f, \tau) = \alpha \hat{\gamma}_Y(f, \tau - 1) G_Y(f, \tau - 1)^2 + (1 - \alpha) P[\hat{\gamma}_Y(f, \tau) - 1], \quad (13)$$

where  $G_Y(f, \tau)$  is the spectral gain function of the multichannel MMSE STSA estimator and is defined as

$$G_Y(f, \tau) = \Gamma(1.5) \frac{\sqrt{\nu_Y(f, \tau)}}{\hat{\gamma}_Y(f, \tau)} \exp(\nu_Y(f, \tau)/2) \cdot [1 + \nu_Y(f, \tau)] I_0(\nu_Y(f, \tau)/2) + \nu_Y(f, \tau) I_1(\nu_Y(f, \tau)/2), \quad (14)$$

$$\nu_Y(f, \tau) = \hat{\xi}_Y(f, \tau) \hat{\gamma}_Y(f, \tau) \{1 + \hat{\xi}_Y(f, \tau)\}^{-1}. \quad (15)$$

Finally, the resultant equi-binaural optimal spectral gain without approximation that can strictly minimize (7) is given by

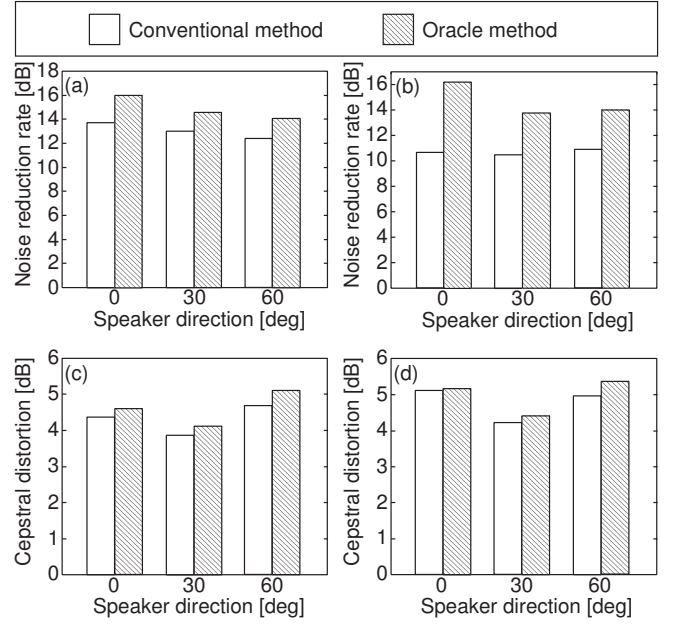
$$G_{\text{oracle}}(f, \tau) = \frac{|h_L(f)||x_L(f, \tau)| + |h_R(f)||x_R(f, \tau)|}{|x_L(f, \tau)|^2 + |x_R(f, \tau)|^2} \cdot G_Y(f, \tau) Y(f, \tau). \quad (16)$$

## 4.2. Experimental Evaluation of Proposed Method I

To evaluate the noise reduction performance of the oracle method, we conducted an experiment on noise reduction. In this experiment, the conventional method and oracle method were compared.

We used 20 utterances (10 males and 10 females from the Japanese newspaper dictation database) as target speech signals and two types of noise signals, namely, white Gaussian noise or speech noise, with spatially diffuse property. Furthermore, the binaural speech signals from three horizontal directions, 0, 30, and 60 degrees, were obtained by convolution of the target speech signals and the HRTF of each direction, where the reverberation time was 200 ms. The test data were obtained by combining the binaural speech signals and noise signals. All signals used in this experiment were 16-kHz-sampled signals. The input SNR was set to 0 dB. The weighting factor  $\alpha$  of the decision-directed estimation was 0.97. The number of iterations in ICA was 300, and the DFT size was 1024. To compare the amount of noise reduction and sound quality, we calculated the noise reduction rate (NRR) [7] (output SNR - input SNR in dB) and cepstral distortion (CD) [9] (a measure of the degree of spectral envelope distortion) of the processed signals.

Figure 2 shows the results for the average NRR and CD of all the target speakers for each direction. While the conventional and oracle methods show almost the same performance for speech distortion, the oracle method can achieve higher noise reduction performance than the conventional method. This indicates the advantage of Proposed Method I.



**Fig. 2.** Experimental results of NRR for (a) white Gaussian noise and (b) speech noise, and CD for (c) white Gaussian noise and (d) speech noise in Proposed Method I.

## 5. PROPOSED METHOD II: GENERALIZED MMSE STSA ESTIMATOR WITH SPEECH PRIOR ESTIMATION

### 5.1. Problem of Oracle Method and Motivation

In the case of using the multichannel MMSE STSA estimator, the noise reduction performance was markedly improved as shown in the previous section. This indicates that the noise reduction performance will be improved by using binaural cues such as the HRTF. However, it is difficult to estimate the HRTF blindly in practice because the shape of the head of hearing-aid users greatly varies. Therefore, we propose a new method that applies different speech priors for the left and right channels as binaural-cue-like information that can be estimated blindly. More specifically, the original MMSE STSA estimator is replaced with a *generalized MMSE STSA estimator* [10, 11]. When the target speaker is located on the right-hand side, the observed signal at the right channel arrives directly, but that of the left channel is more reverberant because of the head diffraction and room reverberation. Hence, the probability density functions (p.d.f.s) of the speech signals observed at each channel vary depending on the speaker direction. However, the original MMSE STSA estimator applied only one fixed speech prior for the target speech signal. Therefore, we apply different speech priors for the left and right channels by replacing the original MMSE STSA estimator with the generalized MMSE STSA estimator. In addition, we introduce a blind speech prior identification algorithm in the next subsection. Figure 3 shows a block diagram of the proposed method.

### 5.2. Parametric Model for Speech

The original (Ephraim's) MMSE STSA estimator assumes that the speech signal obeys a Gaussian distribution. How-

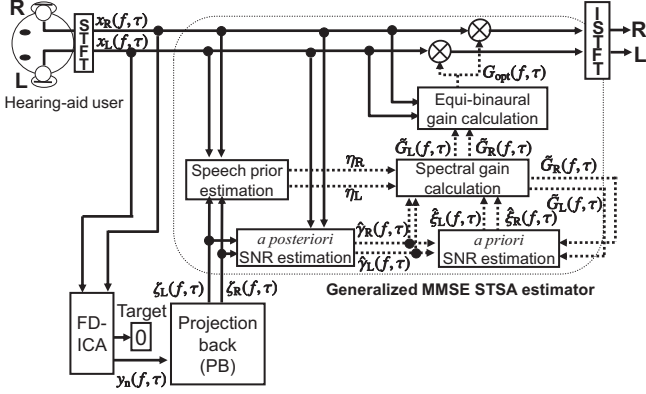


Fig. 3. Block diagram of Proposed Method II.

ever, it is well known that the speech signal has a more spiky p.d.f. similar to a Laplacian distribution. Therefore, in the generalized MMSE STSA estimator, the *generalized gamma distribution* is utilized to model the amplitude spectral grid signal in the time-frequency domain [10]. Its p.d.f. is written as

$$p(a) = 2\phi^\eta \Gamma(\eta)^{-1} a^{2\eta-1} \exp(-\phi a^2), \quad (17)$$

where  $\phi = \eta/E\{|a|^2\}$  and  $\eta$  ( $0 < \eta < 1$ ) is a shape parameter;  $\eta = 1$  gives a Rayleigh distribution that corresponds to a Gaussian signal, and a smaller value of  $\eta$  corresponds to a super-Gaussian signal.

### 5.3. Generalized MMSE STSA Estimator [10]

In the generalized MMSE STSA estimator, a posteriori and a priori SNRs at each channel can be estimated in a similar way to those in the original MMSE STSA estimator using (2) and (3), respectively. Also, the spectral gain function is defined as

$$\tilde{G}_*(f, \tau) = \frac{\sqrt{\tilde{\nu}_*(f, \tau)} \Gamma(\eta_* + 0.5) \Phi(0.5 - \eta_*, 1, -\tilde{\nu}_*(f, \tau))}{\hat{\gamma}_*(f, \tau) \Gamma(\eta_*) \Phi(1 - \eta_*, 1, -\tilde{\nu}_*(f, \tau))}, \quad (18)$$

$$\tilde{\nu}_* = \hat{\xi}_*(f, \tau) (\eta_* + \hat{\xi}_*(f, \tau))^{-1} \hat{\gamma}_*(f, \tau), \quad (19)$$

where  $\Phi$  is a confluent hypergeometric function. The gain  $\tilde{G}_*(f, \tau)$  includes a shape parameter  $\eta_*$  that should represent the speech p.d.f. prior. In Sects. 5.4 and 5.5, we describe how to blindly estimate  $\eta_*$ .

### 5.4. Shape Parameter and Kurtosis

Regarding the generalized gamma distribution  $p(a)$  in (17), the  $m$ th-order moment can be written as

$$\mu_m(a) = \int_0^\infty a^m p(a) da = \frac{\Gamma(\eta + \frac{m}{2})}{\Gamma(\eta)} \phi^{-\frac{m}{2}}. \quad (20)$$

Then, the kurtosis of the generalized gamma distribution is calculated as

$$\text{kurt} = \mu_4(a) / \mu_2^2(a) = (\eta + 1) / \eta. \quad (21)$$

Furthermore, the shape parameter  $\eta$  is given by

$$\eta = (\text{kurt} - 1)^{-1}. \quad (22)$$

From this relation, the shape parameter of the subjective speech signal can be estimated by obtaining its kurtosis value. In general, however, it is difficult to directly estimate the kurtosis of a speech signal because of its contamination by additive noise. Hereafter, an algorithm for speech kurtosis estimation is described.

### 5.5. Estimation of Speech Kurtosis and Gain Function

Hereafter, we define complex-valued variables of the observed (noisy speech) signal, the original speech signal, and the noise signal of each channel as  $(x_r^L + ix_i^L)$  and  $(x_r^R + ix_i^R)$ ,  $(s_r^L + is_i^L)$  and  $(s_r^R + is_i^R)$ ,  $(n_r^L + in_i^L)$  and  $(n_r^R + in_i^R)$ , respectively, where  $x_r = s_r + n_r$  and  $x_i = s_i + n_i$ . The subscripts  $r$  and  $i$  respectively represent the real and imaginary parts of signals. Only the statistics of  $(x_r^L + ix_i^L)$ ,  $(x_r^R + ix_i^R)$ ,  $(n_r^L + in_i^L)$ , and  $(n_r^R + in_i^R)$  are observable, whereas those of  $(s_r^L + is_i^L)$  and  $(s_r^R + is_i^R)$  are unknown values to be estimated.

We can estimate the resultant kurtosis of the speech amplitude spectrum as

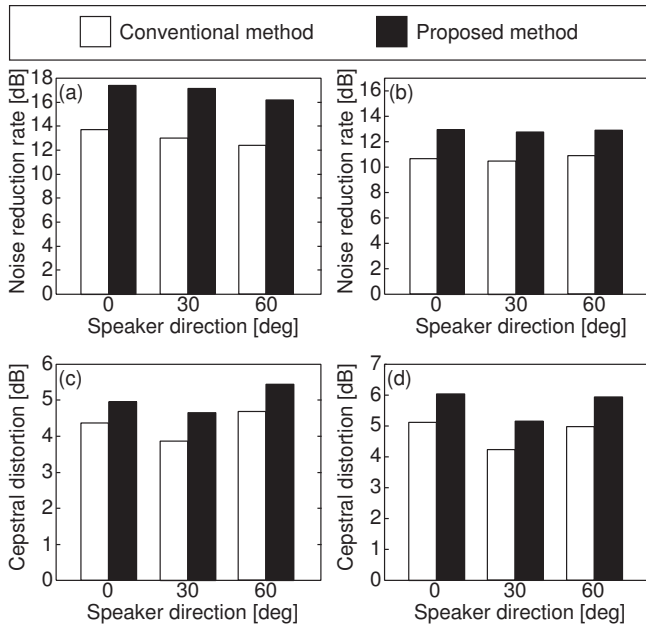
$$\begin{aligned} \text{kurt}_{\text{sp}}^* = & \left[ \mu_4(x_r^*) + \mu_4(x_i^*) - \mu_4(n_r^*) - \mu_4(n_i^*) \right. \\ & + 6\mu_2^2(n_r^*) + 6\mu_2^2(n_i^*) \\ & + 2\mu_2(x_r^*)\mu_2(x_i^*) + 2\mu_2(n_r^*)\mu_2(n_i^*) \\ & - 6\mu_2(x_r^*)\mu_2(n_r^*) - 6\mu_2(x_i^*)\mu_2(n_i^*) \\ & - 2\mu_2(x_r^*)\mu_2(n_i^*) - 2\mu_2(x_i^*)\mu_2(n_r^*) \left. \right] \\ & \cdot \left[ \mu_2^2(x_r^*) + \mu_2^2(x_i^*) + \mu_2^2(n_r^*) + \mu_2^2(n_i^*) \right. \\ & + 2\mu_2(x_r^*)\mu_2(x_i^*) - 2\mu_2(x_r^*)\mu_2(n_r^*) \\ & - 2\mu_2(x_r^*)\mu_2(n_i^*) - 2\mu_2(x_i^*)\mu_2(n_r^*) \\ & \left. - 2\mu_2(x_i^*)\mu_2(n_i^*) + 2\mu_2(n_r^*)\mu_2(n_i^*) \right]^{-1}. \quad (23) \end{aligned}$$

For the detailed derivation of (23), see Ref. [12].

The shape parameter of the speech p.d.f. at each channel can be estimated using these kurtosis and (22). Finally, the equi-binaural optimal spectral gain estimated by the proposed method is obtained by inserting (18) into (9), as

$$\begin{aligned} \tilde{G}_{\text{opt}}(f, \tau) & = \frac{|x_L(f, \tau)|^2 \sqrt{\tilde{\nu}_L(f, \tau)} \Gamma((\text{kurt}_{\text{sp}}^L - 1)^{-1} + 0.5)}{\{|x_L(f, \tau)|^2 + |x_R(f, \tau)|^2\} \hat{\gamma}_L(f, \tau) \Gamma((\text{kurt}_{\text{sp}}^L - 1)^{-1})} \\ & \cdot \frac{\Phi(0.5 - (\text{kurt}_{\text{sp}}^L - 1)^{-1}, 1, -\tilde{\nu}_L(f, \tau))}{\Phi(1 - (\text{kurt}_{\text{sp}}^L - 1)^{-1}, 1, -\tilde{\nu}_L(f, \tau))} \\ & + \frac{|x_R(f, \tau)|^2 \sqrt{\tilde{\nu}_R(f, \tau)} \Gamma((\text{kurt}_{\text{sp}}^R - 1)^{-1} + 0.5)}{\{|x_L(f, \tau)|^2 + |x_R(f, \tau)|^2\} \hat{\gamma}_R(f, \tau) \Gamma((\text{kurt}_{\text{sp}}^R - 1)^{-1})} \\ & \cdot \frac{\Phi(0.5 - (\text{kurt}_{\text{sp}}^R - 1)^{-1}, 1, -\tilde{\nu}_R(f, \tau))}{\Phi(1 - (\text{kurt}_{\text{sp}}^R - 1)^{-1}, 1, -\tilde{\nu}_R(f, \tau))}. \quad (24) \end{aligned}$$





**Fig. 4.** Results of NRR for (a) white Gaussian noise and (b) speech noise, and CD for (c) white Gaussian noise and (d) speech noise in Proposed Method II.

The resultant output signals for the left and right ears can be obtained in the same manner as (6), as

$$z_*(f, \tau) = \tilde{G}_{\text{opt}}(f, \tau)x_*(f, \tau). \quad (25)$$

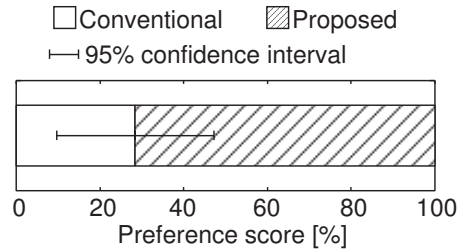
### 5.6. Experimental Evaluation of Proposed Method II

To evaluate Proposed Method II, we conducted an experiment under the same conditions as those of Proposed Method I (see Sect. 4.2). Figure 4 shows the results for the average NRR and CD of all the target speakers for each direction. Although Proposed Method II achieves greater noise reduction than the conventional method, it leads to more speech distortion. Therefore, a trade-off exists between the amount of noise reduction and speech distortion in the conventional and proposed methods.

Since we found the above-mentioned trade-off, we next conducted a subjective preference test for settling the performance competition, focusing on the human impression of the enhanced speech. The result of the preference test is shown in Fig. 5. Nine examinees participated in the preference test. A pair of signals processed using the conventional method and Proposed Method II, for which the type of noise and the direction the of target speech were selected randomly, were presented to participants, who were asked to select which signal they preferred. As shown in Fig. 5, Proposed Method II gains a higher preference score than the conventional method. This well indicates the improvement of the proposed method in terms of the human perceptual impression.

## 6. CONCLUSION

In this paper, we addressed some variations of the source-localization-preserved MMSE STSA estimator used for binaural hearing aids. As a practical solution, we proposed a



**Fig. 5.** Result of preference test for Proposed Method II.

generalized MMSE STSA estimator to apply different speech priors for the left and right channels. From an objective evaluation, it was shown that there is a trade-off between NRR and CD in the conventional and proposed methods. However, in the subjective evaluation, the proposed method achieved a higher preference score than the conventional method according to human perception.

## 7. REFERENCES

- [1] T. van den Bogaert et al., "Speech enhancement with multichannel Wiener filter techniques in multimicrophone binaural hearing aids," *Journal of Acoust. Soc. America*, vol.125, no.1, pp.360–371, 2009.
- [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," *IEEE Trans. Audio, Speech and Lang. Process.*, vol.17, no.4, pp.521–533, 2009.
- [3] K. Reindl et al., "Analysis of two generic Wiener filtering for binaural speech enhancement in hearing aids," *Proc. EUSIPCO*, pp.989–993, 2010.
- [4] Y. Ephraim et al., "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust. Speech Signal Process.*, vol.ASSP-32, no.6, pp.1109–1121, 1984.
- [5] P. Comon, "Independent component analysis, a new concept?," *Signal Process.*, vol.36, pp.287–314, 1994.
- [6] H. Saruwatari et al., "Binaural hearing aid using sound-localization-preserved MMSE STSA estimator with ICA-based noise estimation," *Proc. IWAENC*, 2010.
- [7] Y. Takahashi and H. Saruwatari et al., "Blind spatial subtraction array for speech enhancement in noisy environment," *IEEE Trans. Audio, Speech and Lang. Process.*, vol.17, no.4, pp.650–664, 2009.
- [8] R. C. Hendriks et al., "On optimal multichannel mean-square error estimators for speech enhancement," *IEEE Signal Process. Lett.*, vol.16, no.10, pp.885–888, 2009.
- [9] L. Rabiner et al., *Fundamentals of speech recognition*, Upper Saddle River, NJ: Prentice Hall PTR, 1993.
- [10] I. Andrianakis et al., "MMSE speech spectral amplitude estimators with chi and gamma speech priors," *Proc. ICASSP*, pp.III-1068–III-1071, 2006.
- [11] C. Breithaupt et al., "Parameterized MMSE spectral magnitude estimation for the enhancement of noisy speech," *Proc. ICASSP*, vol.1071, pp.4037–4040, 2008.
- [12] R. Wakisaka et al. "Speech prior estimation for generalized minimum mean-square error short-time spectral amplitude estimator." *IEICE Trans. Fundamentals*, vol.95-A, no.2, pp.591–595, 2012.