Two-Stage Binaural Speech Enhancement with Wiener Filter for High-Quality Speech Communication

Junfeng Li[1*], Shuichi Sakamoto[2], Satoshi Hongo[3], Masato Akagi[1], and Yôiti Suzuki[2]

[1] School of Information Science, Japan Advanced Institute of Science and Technology

[2] Research Institute of Electrical Communication, Tohoku University

[3] Department of Design and Computer Applications, Miyagi National College of Technology

Abstract Speech enhancement has been researched extensively for many years to provide high-quality speech communication in the presence of background noise and concurrent interference signals. Human listening is robust against these acoustic interferences using only two ears, but state-of-the-art two-channel algorithms function poorly. Motivated by psychoacoustic studies of binaural hearing (equalization–cancellation (EC) theory), in this paper, we propose a two-stage binaural speech enhancement with Wiener filter (TS-BASE/WF) approach that is a two-input two-output system. In this proposed TS-BASE/WF, interference signals are first estimated by equalizing and cancelling the target signal in a way inspired by the EC theory, a time-variant Wiener filter is then applied to enhance the target signal given the noisy mixture signals. The main advantages of the proposed TS-BASE/WF are (1) effectiveness in dealing with non-stationary multiple-source interference signals, and (2) success in preserving binaural cues after processing. These advantages were confirmed according to the comprehensive objective and subjective evaluations in different acoustical spatial configurations in terms of speech enhancement and binaural cue preservation.

Keywords: Binaural masking level difference; Equalization-cancellation model; Two-stage binaural speech enhancement (TS-BASE); Binaural cue preservation; Sound localization

1 Introduction

Speech is the most natural and important means of human–human communication in our daily life. Speech communication has been an indispensable component of our society [1]. However, this communication is usually hampered because of the presence of background noise and competing interference signals. To provide high-quality speech communication, speech enhancement techniques have been examined actively in the literature [2, 3]. Mo-

tivated by the good selective hearing ability of normal-hearing persons, much research interest has been paid in recent years to develop two-input two-output binaural speech enhancement systems [4].

The last decades have brought marked advancements in speech enhancement and in understanding of the human hearing mechanism in psychoacoustics, usually in a separate way. Various speech enhancement algorithms have been reported in the literature [2, 3] with many promising applications (e.g., telecommunications and hearing assistant systems). Meanwhile, psychoacoustic studies of binaural hearing show that considerable benefits in understanding a signal in noise can be obtained when either the phase or level differences of the signal at the two ears are not the same as those of the maskers, namely *binaural masking level difference* (BMLD) [5]. Moreover, the binaural cues in signals make it possible to localize their sources and give birth to the perceptual impression of the acoustic scene [5]. According to BMLD, it is believed that speech enhancement systems with binaural cue preservation are much preferred because of the additional benefits in speech enhancement and the perceptual impression of the acoustic scene.

Regarding speech enhancement, in comparison with single-channel techniques (e.g., spectral subtraction [6], Wiener filter [7] and statistic model-based estimators [8]), multi-channel techniques have demonstrated great potential in reducing both stationary and non-stationary interference signals because of the spatial filtering capability provided by multiple spatially distributed microphones [3]. Typical multi-channel approaches are *delay-and-sum* beamformer, *generalized sidelobe canceller* (GSC) beamformer [10], *transfer function GSC* (TF-GSC) [11], GSC with post-filtering [12], *multi-channel Wiener filter* (MWF) [13] and *blind source separation* (BSS) [14]. Many of these multi-channel traditional speech enhancement algorithms have been extended from monaural scenarios to binaural scenarios [15, 16, 17, 18, 19, 20]. Zurek *et al.* extended the original GSC beamformer [10] to binaural scenarios for hearing aids [15, 16]. Campbell *et al.* applied a sub-band GSC beamformer to binaural noise reduction [17, 18]. A common problem associated with these approaches is that no process for equalizing the differences in binaural cues of the target or interference signals is explicitly involved. Suzuki *et al.* suggested introduction of the binaural cues into the constraints of adaptive beamformers to perform adaptive beamforming and preserve the binaural cues within a certain range or direction [19]. Recently, Klasen *et al.* extended the monaural MWF algorithm [13] to the binaural scenario to preserve binaural cues without greatly sacrificing noise reduction performance [20]. However, the adaptive MWF beamformer with two microphones is only optimal for canceling a sin-

gle directional interference. A similar problem is also associated with BSS-based binaural systems, for example, the one proposed by Aichner *et al.* [14].

The multi-channel binaural approaches described above generally involve using a large array of spatially distributed microphones to achieve higher spatial selectivity, which suffers from the high computational cost. In recent years, many multi-channel binaural speech enhancement systems have evolved into two-input two-output binaural systems that are characterized by the small physical size and the low computational cost [4]. Dorbecker *et al.* proposed a two-input two-output spectral subtraction approach based on the assumption of zero correlation between noise signals on two microphones [21], which is rarely satisfied in practical environments. Kollmeier *et al.* introduced a binaural noise reduction scheme based on interaural phase difference (IPD) and interaural level difference (ILD) in the frequency domain [22]. This method was further considered by Nakashima *et al.*, who referred to it as *frequency domain binaural model* (FDBM), in which interference suppression is realized by distinguishing the target and interference signals based on estimates of their directions [23]. Lotter *et al.* proposed a dual-channel speech enhancement approach based on superdirective beamforming under the assumption of a diffuse noise field [24]. More recently, we extended the two-microphone noise reduction method that we proposed previously [25] to the two-output scenario [26], which preserves partial binaural cues at outputs under the assumption of the target signal in front.

To account for BMLD, the equalization–cancellation (EC) theory that distinguishes the target and interference signals based on the dissimilarity of their binaural cues has been widely studied in psychoacoustics [27, 28, 29]. Inspired by the EC theory, in this paper, we propose a two-stage binaural speech enhancement with Wiener filter (TS-BASE/WF) approach, which is essentially a two-input two-output system, for high-quality speech communication. In this proposed TS-BASE/WF, interference signals are first estimated by performing equalization and cancellation processes for the target signal inspired by the EC theory, and a time-variant Wiener filter is then applied to enhance the target signal given the noisy mixture signals. The cancellation strategy in the proposed TS-BASE/WF algorithm differs from that in the original realization of the EC theory [28, 29] in which the cancellation is performed for interference signals, and also differs from those used in many existing systems [15, 16, 17, 18, 21, 26] in which no equalization process is performed prior to cancellation. The main advantages of the proposed TS-BASE/WF approach are (1) effectiveness in dealing with non-stationary multiple-source interference signals, and (2) success in preserving binaural cues after processing. Comprehensive experimental

results in various spatial configurations show that the proposed TS-BASE/WF approach can suppress non-stationary multiple interference signals and preserve binaural cues (i.e. sound source localization) in all tested spatial scenarios.

The remainder of this paper is organized as follows. In section 2, the binaural signal model used in the study is described. The proposed TS-BASE/WF approach, which consists of interference estimation through the EC processes for the target signal in the way inspired by the EC theory and target signal enhancement through the Wiener filter, is detailed in section 3. In section 4, comprehensive experiments were conducted to assess the performance of the proposed TS-BASE/WF approach in terms of speech enhancement and binaural cue preservation. Discussion is provided in section 5, followed by conclusion in section 6.

2 Binaural Signal Model

In binaural processing, the signals at the left and right ears differ not only in the *interaural time difference* (ITD), which is produced because it takes longer for the sound to arrive at the ear that is most distant from the source, but also in the *interaural intensity difference* (IID), which is produced because the signal to the ear closer to the source is more intense as a result of the shadowing effect of the head. Moreover, these signals are corrupted by additive interference signals. Consequently, the observed signals, $X_L(k,\ell)$ and $X_R(k,\ell)$, in the kth frequency bin and the ℓth frame at the left and right ears, can be written as

$$X_L(k,\ell) = H_L(k,\ell)S(k,\ell) + N_L(k,\ell) = S_L(k,\ell) + N_L(k,\ell), \tag{1}$$

$$X_R(k,\ell) = H_R(k,\ell)S(k,\ell) + N_R(k,\ell) = S_R(k,\ell) + N_R(k,\ell), \tag{2}$$

where k and ℓ respectively denote the frequency bin index and the frame index; $S_i(k,\ell)$ and $N_i(k,\ell)$, $(i = L, R)$, are the *short-time Fourier transforms* (STFTs) of the target and noise signals; $H_i(k)$, $(i = L, R)$, represents the transfer functions between the target sound source to two ears, referred to as *head-related transfer function* (HRTF) in the context of binaural hearing. The noise signals might be a combination of multiple interference signals and background noise. In this study, the direction of the target signal is assumed to be known *a priori*. However, no restriction is imposed on the number, location and content of the interference noise sources.

3 Two-Stage Binaural Speech Enhancement with Wiener Filter

As one inspired consideration of this study, EC theory was originally suggested by Kock [27] and subsequently developed by Durlach [28, 29]. According to the EC theory, when the subject is presented with a binaural-masking stimulus, the auditory system attempts to eliminate the masking components by transforming the total signal in one ear relative to the total signal in the other ear until the masking components are identical in both ears (equalization process). Then the total signal in one ear is subtracted from the total signal in the other ear (cancellation process) [28, 29].

Many existing binaural speech enhancement algorithms [15, 16, 17, 18, 21, 26] involve the cancellation process without equalization, thus, they fail to cancel the signals with different binaural cues. Inspired by the essential concept of EC theory, in this paper, we propose a two-stage binaural speech enhancement with Wiener filter (TS-BASE/WF) approach, which consists of: (1) interference estimation by equalizing and cancelling the target signal components inspired by the EC theory, followed by a compensation procedure; (2) target signal enhancement by a time-variant Wiener filter. A block diagram of the proposed TS-BASE/WF system is portrayed in Fig. 1.

3.1 Estimation of interference signals

The objective of the first stage of the TS-BASE/WF is to estimate interference signals at two ears by equalizing and cancelling the target components in the input mixtures. The outputs are then further compensated to yield accurate estimates for interference components in the input noisy signals, as shown in Fig. 1.

3.1.1 Equalization and Cancellation of the target signal

In binaural hearing and binaural applications, HRTFs are normally involved to exhibit the differences in amplitude and phase of signals at the left and right ears. To compensate for these differences, the equalization process for the binaural intensity and phase differences must be performed prior to the cancellation process. The cancellation of the target signal is achieved, in this study, by application of the equalization and cancellation processes for the target signal, yielding the interference-only outputs. It is realized specifically in the following two steps.

1. *In the "equalization" (E) process*, two equalizers are applied to the left and right input signals to equalize the target signal components in these inputs. This equalization process compensates for the differences in intensity and phase of the target signal components at the two ears, caused by shadowing effects of the head introduced by HRTFs. Specifically, given the binaural inputs, two equalizers—$\mathbf{W}_L(k,\ell)$ and $\mathbf{W}_R(k,\ell)$—are obtained using the *normalized least mean square* (NLMS) algorithm, which is given as

$$\mathbf{W}_L(\ell+1)=\mathbf{W}_L(\ell)+\mu\frac{\mathbf{X}_L(\ell)}{||\mathbf{X}_L(\ell)||^2}\left[\mathbf{X}_R(\ell)-\mathbf{W}_L^T(\ell)\mathbf{X}_L(\ell)\right], \tag{3}$$

$$\mathbf{W}_R(\ell+1)=\mathbf{W}_R(\ell)+\mu\frac{\mathbf{X}_R(\ell)}{||\mathbf{X}_R(\ell)||^2}\left[\mathbf{X}_L(\ell)-\mathbf{W}_R^T(\ell)\mathbf{X}_R(\ell)\right], \tag{4}$$

where $\mathbf{W}_i(\ell)=\Big[W_i(1,\ell),W_i(2,\ell),\ldots,W_i(K,\ell)\Big]^T$, $\mathbf{X}_i(\ell)=\Big[X_i(1,\ell),X_i(2,\ell),\ldots,X_i(K,\ell)\Big]^T$ $(i=L,R)$. In addition, superscript ^{T} denotes the transpose operator; K stands for the STFT length, and μ is the step size.

Based on the assumption that the arrival direction of the target signal is known *a priori*, the two equalizers are pre-calibrated in this study in the absence of interference signals. Specifically, the binaural input signals generated by convolving a white noise sequence with the corresponding *head-related impulse response* (HRIR) are used as inputs of the NLMS algorithm to calibrate the two equalizers.

2. *In the "cancellation" (C) process*, the coefficients of two equalizers are fixed and applied to the observed mixture signals in the presence of interference signals. Because the equalizers have been calibrated in the scenarios without interference signals, the target components of the equalizer-filtered left (right) channel input signal are expected to be approximately, if not exactly, equivalent to the target components of the right (left) channel input signal. Consequently, the target-cancelled signals are derived by subtracting the equalizer-filtered inputs at one ear from the input signals at the other ear, given as

$$\begin{aligned} Z_L(k,\ell) &= X_L(k,\ell)-W_R(k,\ell)X_R(k,\ell) \\ &\approx N_L(k,\ell)-W_R(k,\ell)N_R(k,\ell), \end{aligned} \tag{5}$$

$$\begin{aligned} Z_R(k,\ell) &= X_R(k,\ell)-W_L(k,\ell)X_L(k,\ell) \\ &\approx N_R(k,\ell)-W_L(k,\ell)N_L(k,\ell). \end{aligned} \tag{6}$$

From Eqs. (5) and (6), it is observed that the target signals are cancelled and the interference-only signals remain.

Although this cancellation strategy originates from the EC theory in psychoacoustics, it differs from the traditional realizations of the EC theory [28, 29]. Traditionally, the E and C processes are performed for interference components, which enables reduction of only one directional interference signal with the two-channel signals at two ears. In practical environments, however, the number of interference signals is usually unknown or infinity (diffuse noise). Thus, the traditional cancellation strategy [28, 29] cannot deal with multiple interference signals and/or diffuse noise in more challenging practical conditions. By performing the E and C processes for the target signal, in contrast, the proposed TS-BASE/WF approach can calculate the interference signals that might include the energy of multiple interference signals and/or diffuse noise, and be further reduced in its second stage. It is because that the number of target signal of interest is usually one at each instant time under practical environments. Consequently, the TS-BASE/WF approach can deal with the problem of multiple interference signals in adverse practical environments.

3.1.2 Compensation for interference signal estimates

Although the EC processes have cancelled the target components and yielded interference-only outputs as shown in Eqs. (5) and (6), the target-cancelled signals differ from the original interference components in the input mixture signals because of the filtering effects introduced by the two equalizers. As a consequence, this problem results in overestimation or underestimation for interference signals, and further endangers a low noise reduction capability or high speech distortion in the second stage of the TS-BASE/WF.

To address this problem, we propose to exploit a time-variant frequency-dependent compensation factor, $C_i(k,\ell)$, to make the target-cancelled signals approximately, if not exactly, equivalent to the interference components in the input mixture signals. This compensation factor $C_i(k,\ell)$ is derived by minimizing the mean square error between the target-cancelled signal and the input mixture signal under the assumption of zero correlation between the target signal and interference signals, formulated as

$$C_i(k,\ell) = \arg\min \mathcal{E}\Big[X_i(k,\ell) - Z_i(k,\ell)C_i(k,\ell)\Big],\ i = L, R, \tag{7}$$

where $\mathcal{E}$ is the expectation operator. The optimal compensation factors can be found by setting the derivatives of the cost functions with respect to the factors $C_i(k,\ell)$ to zeros. Based on Wiener theory, the optimal compensators $C_i^{\text{opt}}(k,\ell)$ are given as

$$C_i^{\text{opt}}(k,\ell) = \frac{\phi_{X_i Z_i}(k,\ell)}{\phi_{Z_i Z_i}(k,\ell)}, \quad i = L, R, \tag{8}$$

where $\phi_{X_iZ_i}(k,\ell)$ denotes the cross-spectral density of $X_i(k,\ell)$ and $Z_i(k,\ell)$, and $\phi_{Z_iZ_i}(k,\ell)$ is the auto-spectral density of $Z_i(k,\ell)$.

Because the interference-only signals after EC processing and the interference components in the input noisy signals come from the same interference sources, the compensation factors $C_i(k,\ell)$ should be dependent on the spatial location of the target signal relative to those of interference signals. Therefore, in practical conditions, in which the sound sources are usually fixed or which move slowly, these compensation factors are much more stationary than other parameters that are based on the power-spectral densities (PSDs) of the signals used in the traditional algorithms [8, 13, 18, 20]. This characteristic provides the proposed TS-BASE/WF approach with high robustness against non-stationary interference.

3.2 Target signal enhancement

For binaural applications, a system that can yield binaural outputs and preserve binaural cues is much preferred. In the proposed TS-BASE/WF, the compensated interference estimates are used to control the gain function of a speech enhancer, which is shared in both left and right channels for binaural cue preservation. In this study, the improved Wiener filter based on the *a priori* SNR is adopted because of its good noise reduction performance and its capability for reducing "musical noise". Its gain function is formulated as [31]

$$G_{WF}(k,\ell) = \frac{\xi(k,\ell)}{1+\xi(k,\ell)}, \tag{9}$$

where $\xi(k,\ell)$ is the *a priori* SNR defined in [8]. With the compensated two-channel interference estimates at two ears, the *a priori* SNR, $\xi(k,\ell)$, is calculated as

$$\xi(k,\ell) = \frac{\mathcal{E}\Big[S_L(k,\ell)S_L^*(k,\ell) + S_R(k,\ell)S_R^*(k,\ell)\Big]}{\mathcal{E}\Big[\Big(C_L(k,\ell)Z_L(k,\ell)\Big)\Big(C_L(k,\ell)Z_L(k,\ell)\Big)^* + \Big(C_R(k,\ell)Z_R(k,\ell)\Big)\Big(C_R(k,\ell)Z_R(k,\ell)\ \Big)^*\Big]}, \tag{10}$$

where the superscript * signifies the conjugative operator. The estimate of the *a priori* SNR, $\xi(k,\ell)$, is updated in a decision-directed scheme, as [8]

$$\xi(k,\ell) = \alpha\frac{\big|S_L(k,\ell-1)\big|^2 + \big|S_R(k,\ell-1)\big|^2}{\mathcal{E}\Big[\big|N_L(k,\ell-1)\big|^2 + \big|N_R(k,\ell-1)\big|^2\Big]} + (1-\alpha)\times\max\Big[\gamma(k,\ell)-1,\ 0\Big], \tag{11}$$

where α $(0<\alpha<1)$ is a forgetting factor and $\gamma(k,\ell)$ is the *a posteriori* SNR, as defined in [8]. This decision-directed estimation mechanism for the *a priori* SNR markedly decreases the residual "musical noise", as detailed in [32].

4 Experiments and Discussion

The performance of the proposed TS-BASE/WF algorithm was examined in one-noise-source and multiple-noise-source conditions, and further compared to that of state-of-the-art two-input two-output binaural speech enhancement algorithms, including two-channel spectral subtraction (TwoChSS) [21], frequency-domain binaural model (FDBM) [22, 23], and two-channel superdirective beamformer (TwoChSDBF) [24]. The parameters used in the implementation of these algorithms were the same as those published. In the implementation of TS-BASE/WF, both frame length and FFT size were set to 64 ms, frame shift was 32 ms, the step size μ used in the NLMS algorithm for calibrating two equalizers was 0.01, and the length of two equalizers was set to 512. Numerous experiments were conducted to evaluate the performance of the tested algorithms extensively, with regard to speech enhancement and binaural cue preservation (i.e. sound localization) in various spatial configurations using both objective and subjective evaluation measures.

4.1 Experimental evaluations for speech enhancement

4.1.1 Experimental configuration

In speech enhancement experiments, 50 continuous speech sentences, in which each utterance was about 3-5 seconds, uttered by three male and two female speakers were randomly selected from NTT database that has a sampling rate of 44.1 kHz at 16 bit resolution [33]. Among these utterances, 10 sentences were used as the target speech signals, the other 40 were used as the interference signals. These signals were then convolved with the HRIRs measured at the MIT Media Laboratory [34] to generate the binaural target and interference signals. The binaural target and interference signals were downsampled to 16 kHz. The interference signals were then scaled to obtain an average input SNR of 0 dB across two channels before being added to the target signals. The binaural noisy input signals were finally generated by adding the scaled binaural interference signals to the binaural target signals.

To examine the efficacy of the studied systems, we performed evaluations in various spatial configurations as listed in Table 1. In Table 1, $S_\theta N_\psi$ denotes the spatial scenario in which the target signal (S) arrives from the direction θ and interference signal(s) (N) come from direction(s) ψ. Directions are defined clockwise with 0° being directly in front of the listener.

4.1.2 Objective evaluations

The improvement in SNR was used to evaluate the speech enhancement performance of the proposed TS-BASE/WF and traditional algorithms objectively. It is defined as

$$\Delta\text{SNR} \quad = \text{SNR}_o - \text{SNR}_i, \tag{12}$$

where SNR_o and SNR_i are the SNRs of the output enhanced signal and the input noisy signal. Actually, the SNR is defined as the ratio of the power of clean speech to that of noise signal embedded in the noisy input signal (SNR_i) or the enhanced signal by the studied algorithms (SNR_o), given as

$$\text{SNR}_i \quad = \quad 10\log_{10}\left(\sum_t s^2(t) \Big/ \sum_t \left[x(t) - s(t)\right]^2\right), \tag{13}$$

$$\text{SNR}_o \quad = \quad 10\log_{10}\left(\sum_t s^2(t) \Big/ \sum_t \left[s(t) - \hat{s}(t)\right]^2\right), \tag{14}$$

where $s(.)$ and $\hat{s}(.)$ are the reference clean speech signal and the enhanced signal processed by the tested algorithms, and $x(.)$ is the noisy input signal. A higher ΔSNR means a higher improvement in speech quality by speech enhancement processing.

Fig. 2 portrays the ΔSNRs averaged across all utterances, as processed using the proposed TS-BASE/WF approach and other traditional algorithms in the one-noise-source conditions S_0N_ψ. The ΔSNR results in more challenging scenarios with multiple noise sources or non-zero arrival directions of the target signal are shown in Fig. 3. All these evaluations were performed separately for signals in the left and right ears.

The ΔSNR results in the one-noise-source conditions presented in Fig. 2 show that all tested algorithms produce positive ΔSNRs (i.e. improved speech quality), and that these ΔSNRs vary greatly with the incoming direction of the interference signal. Specifically, the ΔSNRs are much higher when the interference signal is close to the ear with which the enhanced signal is under evaluation. This is the case in which the input signals are more noisy with low SNRs. As an example, Fig. 2(a) shows that the ΔSNRs at the left ear (with the interference signal at the left side of the head) are much higher than those with the interference signal at the right side. Regarding comparisons of the studied algorithms, the TwoChSDBF and FDBM algorithms yield low ΔSNRs under all tested conditions. The low capability of TwoChSDBF algorithm in speech enhancement is attributed to its assumption of a diffuse noise field [24]. The performance of FDBM algorithm is limited by its low capability of distinguishing the arrival directions of the target and interference signals in low SNR conditions. Comparison with the TwoChSDBF and FDBM algorithms

reveals that the TwoChSS algorithm yields much larger ΔSNRs because of the use of a noise estimation technique based on spatial information [21]. In contrast with all traditional algorithms, the proposed TS-BASE/WF algorithm provides the highest ΔSNRs in all tested conditions, especially when the interference signal is close to the ear under evaluation. The high speech-enhancement performance of the proposed TS-BASE/WF results from its accurate noise estimation capability through the equalization and cancellation processes for the target signal inspired by the EC theory. One observation of interest is that the ΔSNRs produced by all studied algorithms in the S_0N_0 and S_0N_{180} conditions are close to 0 dB because the target signal and interference signals involve equivalent binaural cues. Consequently, all tested algorithms fail to distinguish the target signal and interference signals based on their binaural cues (i.e. spatial information). Similar results are observed for the right ear, as portrayed in Fig. 2(b).

The ΔSNR results shown in Fig. 3 demonstrate that the studied algorithms can enhance the speech quality (i.e. the positive ΔSNRs) at the left and right ears in all multiple-noise-source conditions. In multiple-noise-source environments, the TwoChSDBF algorithm again gives the lowest SNR improvements. Comparatively, the FDBM and TwoChSS systems coequally produce much larger ΔSNRs. The proposed TS-BASE/WF algorithm provides significant improvements in SNR at both left and right ears in the presence of multiple interference sources. Another important observation is that in the conditions with non-zero arrival direction of the target signal (i.e., $S_{90}N_0$, $S_{90}N_{270}$, and $S_{45}N_{315}$), the traditional TwoChSDBF and TwoChSS algorithms show very limited SNR improvements. The FDBM approach gives much higher SNR improvements at the left ear. Regarding results observed at the right ear, the TwoChSS and FDBM algorithms show the markedly decreased ΔSNR in the $S_{90}N_0$ scenario and even the negative ΔSNRs in $S_{90}N_{270}$ and $S_{45}N_{315}$ conditions, and the TwoChSDBF algorithm shows a relative robustness in these conditions. In contrast, the proposed TS-BASE/WF algorithm yields considerable SNR improvements at the left ear (shown in Fig. 3(a)), and small SNR improvements at the right ear (shown in Fig. 3(b)), which are higher than those of the traditional algorithms (except for the TwoChSDBF algorithm in the $S_{90}N_{270}$ condition). The low ΔSNRs at the right ear are attributed to the weak noise components (i.e. high SNRs) there because the target signal is closer to that ear, although the interference signal is more distant.

4.1.3 Subjective evaluations

The performance of the studied algorithms was perceptually assessed further through listening tests. In these evaluations, the processed signals at the left and right ears were presented separately to listeners.

In subjective evaluations, 6 utterances were selected from the NTT database and used as the target speech signals and another 24 different utterances were used as the interfering signals. The noisy mixture signals were generated as described in section 4.1.2 at SNR of 0 dB in the following spatial configurations: S_0N_{60}, S_0N_{3a}, S_0N_{4a}, and $S_{90}N_0$. The resultant 24 (4×6) noisy speech sentences at the left ear were then processed using the four tested algorithms. In each scenario, the processed 24 speech signals, along with the 6 unprocessed noisy signals at the left ear as references, were then presented randomly through a headphone at a comfortable volume in a soundproof room to 10 graduate students with normal hearing capability. The same procedure was also performed for the signals to the right ear. Each listener was instructed to rate the speech quality based on their preference in terms of *mean opinion score* (MOS): 1 = bad, 2 = poor, 3 = fair, 4 = good, 5 = excellent.

The speech enhancement performance of the studied algorithms was evaluated subjectively in terms of the MOS improvement ΔMOS, calculated as

$$\Delta\text{MOS} = \text{MOS}_{enhanced} - \text{MOS}_{unproc}, \tag{15}$$

where MOS_{unproc} and $\text{MOS}_{enhanced}$ are the MOS scores of the unprocessed noisy signal and the enhanced signal by the tested algorithms. A high ΔMOS indicates high improvement in speech quality.

The ΔMOS results of the studied algorithms in different acoustic scenarios are plotted in Fig. 4. Results show that all tested algorithms yield different degrees of MOS improvements at two ears in the tested conditions.

In the conditions with the target signal arriving from 0°, only small improvements in MOS are observed when using the TwoChSDBF algorithm. In comparison with the TwoChSDBF algorithm, the TwoChSS algorithm provides much larger ΔMOS in these conditions. Based on the interaural information of the binaural inputs, the FDBM algorithm shows robust MOS improvements as the number of interference signals increases. Furthermore, the proposed TS-BASE/WF algorithm offers the largest ΔMOS (i.e. the highest speech quality) among the tested algorithms in all spatial configurations. These MOS improvements at the two ears show only a slight decrease with the increasing number

of interference signals. The perceptual preference of the enhanced signals using the proposed TS-BASE/WF is also attributed to the marked reduction of "musical noise" [32], while the traditional algorithms are inefficient in dealing with "musical noise".

More importantly, in the acoustic condition $S_{90}N_0$, the traditional TwoChSS method does not function well because it normally assumes that the target signal comes from 0°. The MOS improvements of the TwoChSDBF algorithm are also limited because of the unreasonable noise field assumption. The FDBM algorithm yields high ΔMOSs by steering the interested direction to the target source. The proposed TS-BASE/WF algorithm exhibits the largest ΔMOSs at both ears by exploiting the direction information of the target signal.

4.2 Experimental evaluations for binaural cue preservation

For binaural processing, in addition to reducing interference components, the capability of preserving binaural cues is another important issue to evaluate. In this subsection, the proposed TS-BASE/WF algorithm is examined with regard to binaural cue preservation (i.e. sound source localization), and further compared with the traditional binaural speech enhancement algorithms used in the preceding section.

4.2.1 Objective evaluations

In objective evaluations for binaural cue preservation, the same target and interference signals as those used in the objective evaluations for speech enhancement were used. The noisy binaural signals were generated with a SNR of 0 dB under spatial configurations: the one-noise-source conditions ($S_{0:30:360}N_0$), and the three-noise-source conditions ($S_{0:30:360}N_{90,180,270}$), where the target source was simulated to be placed around the listener at positions from 0° to 360° in increments of 30°, and the interfering signal(s) were placed at fixed position(s).

4.2.1.1 Objective evaluation measures

The respective efficacies of the proposed TS-BASE/WF and other traditional algorithms in binaural cue preservation were evaluated objectively using the ITD error (E_{ITD}) and the ILD error (E_{ILD}) of the outputs.

The ITD error (E_{ITD}) is defined as [30]

$$E_{ITD} = \frac{\left|\angle c_{enhanced} - \angle c_{clean}\right|}{\pi}, \tag{16}$$

where $\angle c_{enhanced}$ and $\angle c_{clean}$ are the phases of the cross spectra (i.e. the approximate ITD estimates) for the enhanced signals $\hat{S}_i$ and clean signals S_i, calculated as (k and ℓ are omitted hereinafter for notational simplicity.)

$$c_{enhanced} = \mathcal{E}\{\hat{S}_L \hat{S}_R^*\}, \quad c_{clean} = \mathcal{E}\{S_L S_R^*\}. \tag{17}$$

In the evaluations, the estimation of the ITD error was only performed in the frequency regions below 2 kHz, since only ITD cues contained in the low-frequency regions are used to localized sounds horizontally for human [5].

Similarly, the ILD error (E_{ILD}) is defined as [30]

$$\mathrm{E}_{ILD} = \left| 10\log_{10} P_{enhanced} - 10\log_{10} P_{clean} \right|, \tag{18}$$

where $P_{enhanced}$ and P_{clean} respectively represent the power ratios (i.e. the approximate ILD estimates) for the enhanced signals and the clean signals, calculated as

$$P_{enhanced} = \frac{\mathcal{E}\{|\hat{S}_L|^2\}}{\mathcal{E}\{|\hat{S}_R|^2\}}, \quad P_{clean} = \frac{\mathcal{E}\{|S_L|^2\}}{\mathcal{E}\{|S_R|^2\}}. \tag{19}$$

The smaller E_{ITD} and E_{ILD} are, the higher the performance of the tested algorithm in binaural cue preservation is.

4.2.1.2 Objective evaluation results

The results in E_{ITD} and E_{ILD} averaged across all tested utterances under the one-noise-source and three-noise-source conditions are shown respectively in Fig. 5 and Fig. 6.

From Fig. 5(a), symmetry of E_{ITD} along with the median plane in the one-noise-source conditions is observed. Two facts contribute to this symmetric property: (1) symmetry of the HRIRs against the median plane; (2) operations in the spectral amplitude/power domain of the studied algorithms. The symmetry of the HRIRs [34] means that the binaural signals from the sources localized at the median plane involve the equivalent binaural cues. Consequently, the binaural cues of the target signal equal those of the interference signals in the $\mathrm{S}_0\mathrm{N}_0$, $\mathrm{S}_{180}\mathrm{N}_0$ and $\mathrm{S}_{360}\mathrm{N}_0$ scenarios. In these cases, all tested algorithms fail to address the interference signal and yield no benefit in reducing E_{ITD}. In other cases in which the target signal is not on the median plane, the operations with real-gain filtering in all tested algorithms result in the symmetric E_{ITD} because their performance depends only on the relative differences of the arrival directions of the target and interference signals. Regarding the comparisons of the studied algorithms, Fig. 5(a) illustrates that all studied algorithms exhibit different degrees of E_{ITD} under one-noise-source conditions. The traditional TwoChSS algorithm yields largest E_{ITD} after processing, which

results from independent processing in two channels. The other traditional algorithms (i.e., TwoChSDBF and FDBM) introduce smaller E_{ITD} for the target signals with different arrival directions. These benefits are provided by the shared use of one filter with a real-value gain function at the left and right ears. The proposed TS-BASE/WF approach shows the smallest E_{ITD} under all tested spatial configurations. This virtue of the TS-BASE/WF algorithm can be attributed to: (1) the shared use of one filter in two channels; (2) its high noise reduction performance. The first factor enables preservation of the ITD cues of the binaural noisy input signals, and the second one significantly decreases the effects of interference components on the preserved ITD cues. Consequently, the proposed TS-BASE/WF algorithm is able to reduce ITD errors considerably in the tested one-noise-source conditions.

The results in the three-noise-source conditions shown in Fig. 5(b) show that the traditional algorithms (TwoChSS, TwoChSDBF, and FDBM) again provide large E_{ITD}. Among the tested algorithms, the proposed TS-BASE/WF provides the smallest E_{ITD} in all tested conditions. Unlike the results shown in Fig. 5(a), these E_{ITD} results under three-noise-source conditions do not demonstrate the perfect symmetry characteristic against the median plane because different interference signals were used, although they are placed at the symmetric 90° and 270° positions.

The results in E_{ILD} under one-noise-source and three-noise-source conditions are shown in Fig 6. Based on these results, it is observed that the TwoChSS algorithm shows the largest E_{ILD} in both one-noise-source and three-noise-source conditions because of the separate processing of binaural input signals. The traditional TwoChSDBF and FDBM algorithms demonstrate still high E_{ILD} in these conditions. The proposed TS-BASE/WF approach markedly reduces the ILD errors (i.e. the lowest E_{ILD}) due to the shared use of one filter in two channels and the high noise reduction capability. Moreover, similar to discussions related to the E_{ITD} results in Fig. 5(a), all of the studied algorithms exhibit symmetric E_{ILD} along the median plane in one-noise-source conditions; non-perfect symmetric characteristics of E_{ILD} are observed in three-noise-source conditions.

Based on the results presented in Figs. 5 and 6, the proposed TS-BASE/WF algorithm offers the lowest ITD and ILD errors (i.e. preserves the binaural cues), which is expected to enable listeners to localize sound sources more accurately and help them to preserve the perceptual impression of the auditory scene.

4.2.2 Subjective evaluations

The objective evaluations presented in section 4.2.1 have proved that the proposed TS-BASE/WF introduces the lowest ITD and ILD errors compared with the traditional algorithms. Therefore, only the proposed TS-BASE/WF algorithm was evaluated to confirm its capability in sound localization perceptually further through listening tests in this subsection.

In the evaluations, the same target and interference signals as those used in the subjective speech enhancement experiments described in section 4.1.3 were used. The binaural input signals were generated at the SNR of 0 dB under the same spatial configurations as those for binaural-cue preservation experiments described in section 4.2.1, and then processed using the proposed TS-BASE/WF algorithm. The resultant 6 binaural enhanced signals were presented randomly to 10 listeners, who also participated in the subjective speech enhancement experiments, through headphones in a soundproof room. Each listener was first pre-trained using the binaural clean signals given the "real" arrival directions of the target clean signals in the absence of interference signals. Subsequently, the listeners participated in the testing procedure: the processed signals were presented randomly. Each listener was then instructed to give one response for the perceived direction of each processed signal. In all, 720 responses (6 utterances $\times$ 10 listeners $\times$ 12 spatial configurations) were used in each noise condition.

The localization results in one-noise-source and three-noise-source conditions are presented in Fig. 7. The area of each circle is proportional to the number of responses. In all, there are 60 (6 utterances $\times$ 10 listeners) responses under each spatial configuration. The ordinate of each panel is the perceived direction, and the abscissa is the "real" direction of the target signal. Fig. 7 shows that the responses are distributed along a diagonal line: the perceived directions are closely consistent with the "real" ones. The front-back confusion is observed in both one-noise-source and three-noise-source conditions. Further observation reveals that when the target signal is in the front and rear regions (around 0° and 180°), most listeners can perceive the correct target directions (except for the front-back confusion). In the lateral area (90° and 270°), the perceived directions are dispersed around the "real" directions. Similar observations were reported for binaural clean signals in an earlier study [5]. In comparison with the results in these two spatial conditions, the variances of the perceived directions for the target signals in one-noise-source condition are slightly lower than those in three-noise-source conditions.

In summary, the objective and subjective evaluations described above confirm that

the proposed TS-BASE/WF algorithm can preserve the binaural cues of the processed target signal, and localize the target sound source after processing in complex acoustical environments, which enables preservation of the perceptual impressions of auditory scenes.

5 Discussion

The cancellation strategy for the target signal in the proposed TS-BASE/WF system differs from that used in the state-of-the-art multi-channel binaural speech enhancement methods [15, 16, 17, 18, 21, 26]. In these traditional methods, no equalization process is performed prior to cancellation. Therefore, the signal to be cancelled is normally assumed with the same binaural cues at the left and right ears, i.e. the sound source is in front. Inspired by the EC theory, on the other hand, the strategy in the TS-BASE/WF involves the equalization process before cancellation. Through performance of the E and C processes, this strategy can cancel the signal placed at arbitrary spatial locations with different binaural cues. In this sense, the proposed cancellation strategy can be regarded as an extension of the traditional cancellation approach.

Although a similar cancellation strategy was also exploited in the systems in [11, 12], the purpose of these traditional systems was merely to suppress interference signals, finally yielding the monaural enhanced target signal that helps to improve the performance of speech recognizers [11, 12]. Regarding high-quality speech communication in binaural scenarios, in addition to speech enhancement, the proposed TS-BASE/WF system gives due attention to preserving the binaural cues that give birth to perceptual impressions of acoustic scenes. Moreover, subtractive-type processing and the binary mask filtering in these traditional systems [11, 12] introduce the annoying "musical noise". On the other hand, the improved Wiener filter based on the *a priori* SNR used in the proposed TS-BASE greatly reduces "musical noise" and improves the quality of the enhanced signal, as reported by listeners in subjective speech enhancement evaluations.

In comparison with the state-of-the-art binaural speech enhancement algorithms tested in section 4, methodologically, the proposed TS-BASE/WF approach, in which the interference signals are first estimated by equalizing and canceling the target signal followed by target signal enhancement, provides high capability in reducing non-stationary multiple interference signals, as shown in section 4.1. Furthermore, the shared use of one filter with real-value gain in two channels enables the proposed TS-BASE/WF to preserve the binaural cues of the noisy input signals. The effects of interference signals on the preserved binaural cues are reduced markedly by the high noise-reduction performance of the TS-

BASE/WF algorithm. Consequently, the proposed TS-BASE/WF approach can preserve binaural cues of the target signal at the binaural outputs, as presented in section 4.2.

6 Conclusion

In this paper, we proposed a two-stage binaural speech enhancement with Wiener filter (TS-BASE/WF) approach inspired by the equalization-cancellation (EC) theory for high-quality speech communication. In the TS-BASE/WF approach, interference signals are first calculated by equalizing and cancelling the target signal inspired by the EC theory, followed by an interference compensation process, and the target signal is then enhanced by the time-variant Wiener filter. The effectiveness of the proposed TS-BASE/WF algorithm in suppressing multiple interference signals was proved by objective SNR improvements and subjective MOS evaluations. The abilities of the proposed TS-BASE/WF algorithm in preserving the binaural cues and sound localization were also confirmed through objective evaluations using binaural cue errors and subjective sound localization experiments.

In the proposed TS-BASE/WF algorithm, the arrival direction of the target signal is assumed to be known *a priori*. This assumption is sometimes not satisfied in some real applications. A future direction for this study is to integrate the direction estimation technique for the target signal. Moreover, the proposed TS-BASE/WF developed in this paper was designed to address multiple interference signals. In real environments, for example in a room, reverberation is another important factor degrading the quality of speech communication. Therefore, we also plan to extend this TS-BASE/WF algorithm to deal jointly with both interference noise signals and reverberation in future research.

References

[1] A. Waibel, "Speech Processing in Support of Human-Human Communication," in Second International Symposium on Universal Communication, pp. 11, Osaka, Japan, 2008.

[2] P. C. Loizou, Eds., "Speech Enhancement: Theory and Practice," CRC Press, 2007.

[3] M. S. Brandstein and D. B. Ward, Eds., "Microphone Arrays: Signal Processing Techniques and Applications," Springer-Verlag, Berlin, 2001.

[4] D.L. Wang and G.J. Brown, "Computational Auditory Scene Analysis: Principles, Algorithms and Applications," Wiley/IEEE Press, 2006.

[5] J. Blauert, "Spatial Hearing: The Psychophysics of Human Sound Localization," Revised Edition, MIT Press, Cambridge, Massachusetts, USA, 1997.

[6] S.F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. on Acoust. Speech Signal Process.*, vol. 27, no. 4, pp. 113-120, 1979.

[7] N. Wiener, "Extrapolation, Interpolation, and Smoothing of Stationary Time Series," New York: Wiley, 1949.

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

[9] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech and Audio Processing*, vol. 3, no. 7, pp. 251-266, 1995.

[10] J. Griffiths, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas Propagat.*, vol. 30, pp. 27-34, 1982.

[11] S. Gannot, D. Burshtein and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Trans. on Signal Processing*, vol. 49, no. 8, pp. 1614-1626, 2001.

[12] N. Roman, S. Srinivasan and D. Wang, "Binaural segregation in multisource reverberant environments," *Journal of the Acoustical Society of America*, vol. 120, no. 6, pp. 4040-4050, 2006.

[13] S. Doclo, A. Spriet, J. Wouters and M. Moonen, "Frequency-Domain Criterion for the Speech Distortion Weighted Multichannel Wiener Filter for Robust Noise Reduction," *Speech Communication*, vol. 49, no. 7-8, pp. 636-656, 2007.

[14] R. Aichner, H. Buchner, M. Zourub, W. Kellermann, "Multi-channel source separation preserving spatial information," in *Proc. ICASSP2007*, pp. I.5-8, 2007.

[15] J.G. Desloge, W.M. Rabinowitz and P.M. Zurek, "Microphone-array hearing aids with binaural output. I: Fixed-processing systems," *IEEE Trans. Speech Audio Processing, vol. 5, no. 6, pp. 529-542,* Nov. 1997.

[16] D.P. Welker, J.E. Greenberg, J.G. Desloge, P.M. Zurek, "Microphone-array hearing aids with binaural output. II: A two-microphone adaptive system," *IEEE Trans. Speech and Audio Processing, vol. 5. no. 6, pp. 543-551,* Nov., 1997.

[17] P.W. Shields and D.R. Campbell, "Improvements in intelligibility of noisy reverberant speech using a binaural subband adaptive noise-cancellation processing scheme," *Journal of the Acoustical Society of America,* vol. 110, no. 6, pp. 3232-3242, 2001.

[18] D. Campbell and P. Shields, "Speech enhancement using sub-band adaptive Griffiths–Jim signal processing," *Speech Communication*, vol. 39, pp. 97-110, 2003.

[19] Y. Suzuki, S. Tsukui, F. Asano, R. Nishimura and T. Sone, "New design method of a binaural microphone array using multiple constraints," *IEICE Trans. Fundamentals*, vol. E82-A, no. 4, pp. 588-595, 1999.

[20] T.J. Klasen, T. Van den Bogaert, M. Moonen, J. Wouters, "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," *IEEE Trans. on Signal Processing*, vol. 55, no. 4, pp. 1579-1585, 2007.

[21] M. Dorbecker, S. Ernst, "Combination of two-channel spectral subtraction and adaptive Wiener post-filtering for noise reduction and dereverberation," EUSIPCO1996, pp. 995-998, 1996.

[22] B. Kollmeier, J. Peissig, V. Hohmann, "Binaural noise-reduction hearing aid scheme with real-time processing in the frequency domain," *Scand. Audiol. Suppl.* vol. 38, pp. 28-38, 1993.

[23] H. Nakashima, Y. Chisaki, T. Usagawa and M. Ebata, "Frequency domain binaural model based on interaural phase and level differences," *Acoustical Science and Technology*, vol. 24, no. 4, pp. 172-178, 2003.

[24] T. Lotter, B. Sauert and Peter Vary, "A stereo input–output superdirective beamformer for dual channel noise reduction," *In Proc., Eurospeech2005*, pp. 2285-2288, 2005.

[25] J. Li, M. Akagi and Y. Suzuki, "A two-microphone noise reduction method in highly non-stationary multiple-noise-source environments," *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Science*, vol. E91-A, no. 6, pp. 1337-1346, 2008.

[26] J. Li, M. Akagi and Y. Suzuki, "Extension of the two-microphone noise reduction method for binaural hearing aids," in *Proc. International Conference on Audio, Language and Image Processing*, pp. 97-101, Shanghai, China, 2008.

[27] W.E. Kock, "Binaural localization and masking," *Journal of the Acoustical Society of America,* vol. 22, pp. 801-804, 1950.

[28] N.I. Durlach, "Equalization and cancellation theory of binaural masking level differences," *Journal of the Acoustical Society of America,* vol. 35, no. 8, pp. 1206-1218, 1963.

[29] N.I. Durlach, "Binaural signal detection: Equalization and cancellation," In J.V. Tobias, editor, *Foundations of Modern Auditory Theory,* vol. 2, pp. 369-462, Academic Press, New York, 1972.

[30] T.V.Bogaert, J. Wouters, S. Doclo and M. Moonen, "Binaural cue preservation for hearing aids using an interaural transfer function multichannel Wiener filter," ICASSP2007, pp. IV565-568, 2007.

[31] P. Scalart and J. Vieira Filho, "Speech enhancement based on a priori signal to noise estimation," *IEEE International Conference Acoustics, Speech, Signal Processing*, vol. 2, pp. 629-632, Atlanta, USA, 1996.

[32] O. Cappe, "Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor," *IEEE Trans. Speech, and Audio Processing,* vol. 2, no. 2, pp. 345-349, 1994.

[33] http://www.ntt-at.com/products_e/speech2002/.

[34] http://sound.media.mit.edu/KEMAR.html.

Table 1: List of spatial scenarios, $S_\theta N_\psi$, under which the speech enhancement capability of the studied algorithms were evaluated. Here, θ represents the arrival direction of the speech source S, ψ represents the arrival direction(s) of the noise source(s).

Scenario	Spatial Scenarios	Description
One-noise-source	S_0N_ψ	speech source at 0°; ψ between 0° and 330°
	$S_{45}N_{315}$	speech source at 45°; noise source at 315°
	$S_{90}N_0$	speech source at 90°; noise source at 0°
	$S_{90}N_{270}$	speech source at 90°; noise source at 270°
Two-noise-source	S_0N_{2a}	noise sources at 60°, 300°
	S_0N_{2b}	noise sources at 120°, 240°
	S_0N_{2c}	noise sources at 90°, 270°
Three-noise-source	S_0N_{3a}	noise sources at 90°, 180°, 270°
	S_0N_{3b}	noise sources at 30°, 60°, 300°
Four-noise-source	S_0N_{4a}	noise sources at 60°, 120°, 180°, 270°
	S_0N_{4b}	noise sources at 45°, 135°, 225°, 315°

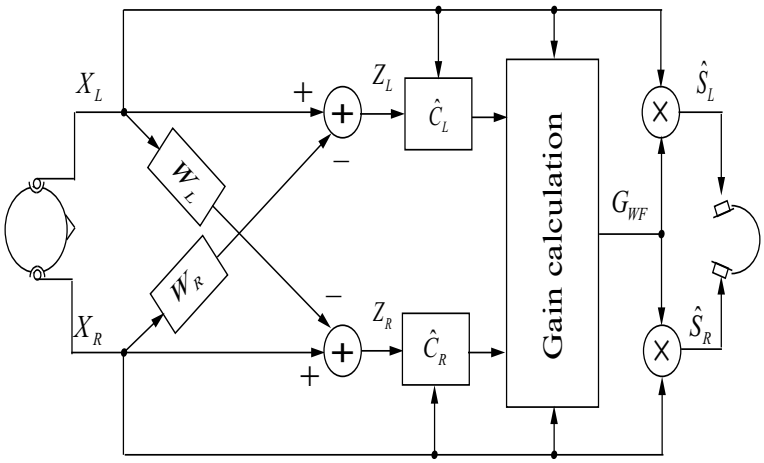


Figure 1: Block diagram of the proposed TS-BASE/WF system.

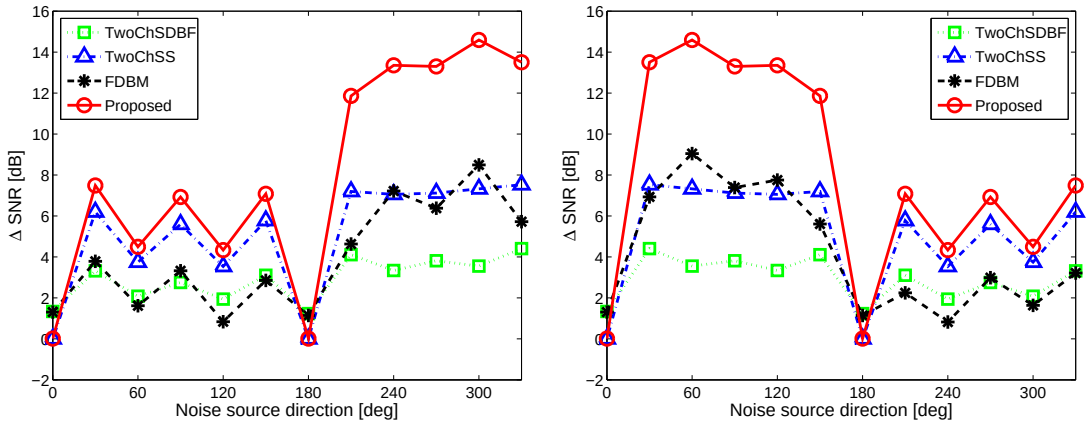
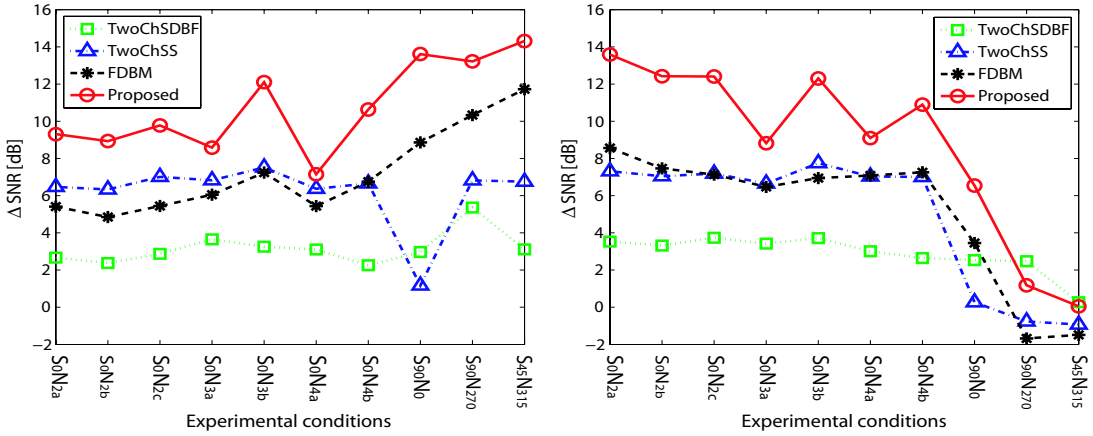
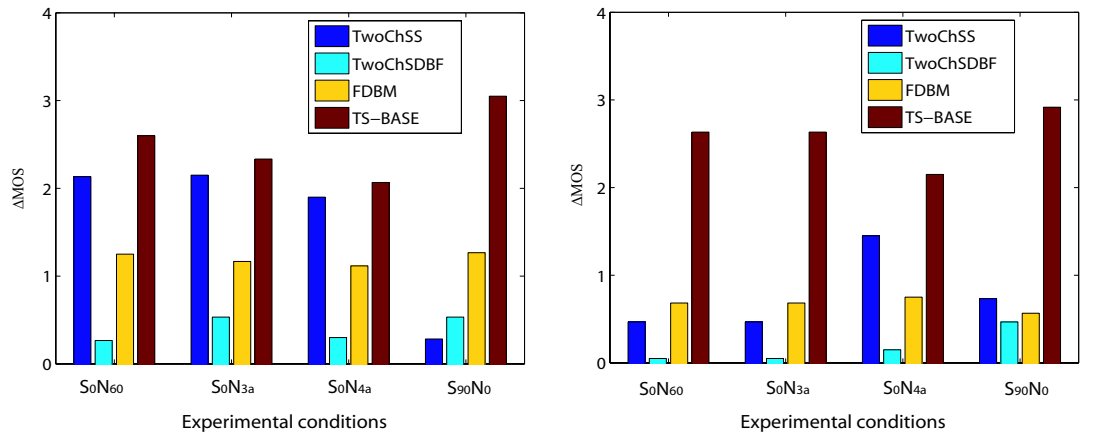


Figure 2: SNR improvements (ΔSNRs) at the left ear (a) and the right ear (b) in one-noise-source conditions, S_0N_ψ, where the speech source is placed at 0° and the interfering signal is at the position from 0° to 330° in increments of 30°.



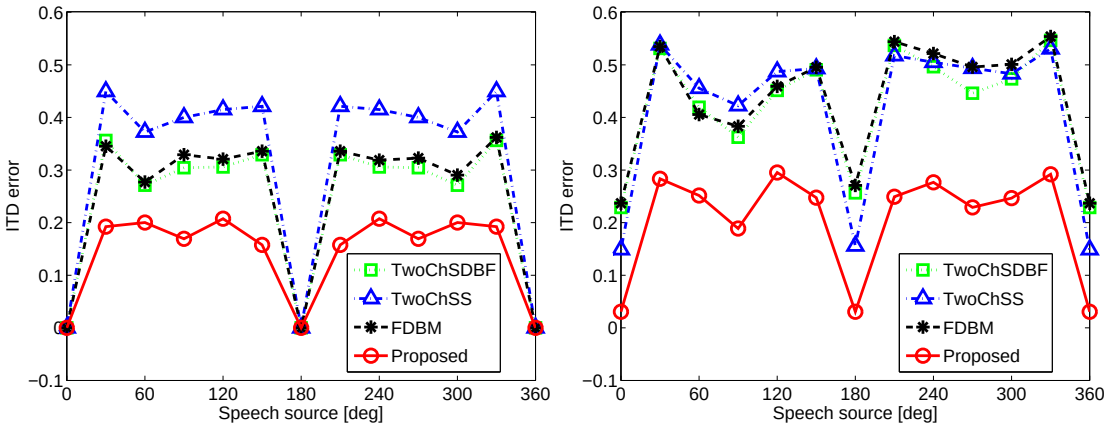
(a) (b)

Figure 3: SNR improvements (ΔSNRs) at the left ear (a) and the right ear(b) in multiple-noise-source conditions, and conditions with non-zero incoming direction of the target signal. The acoustical spatial configurations are presented in Table 1.



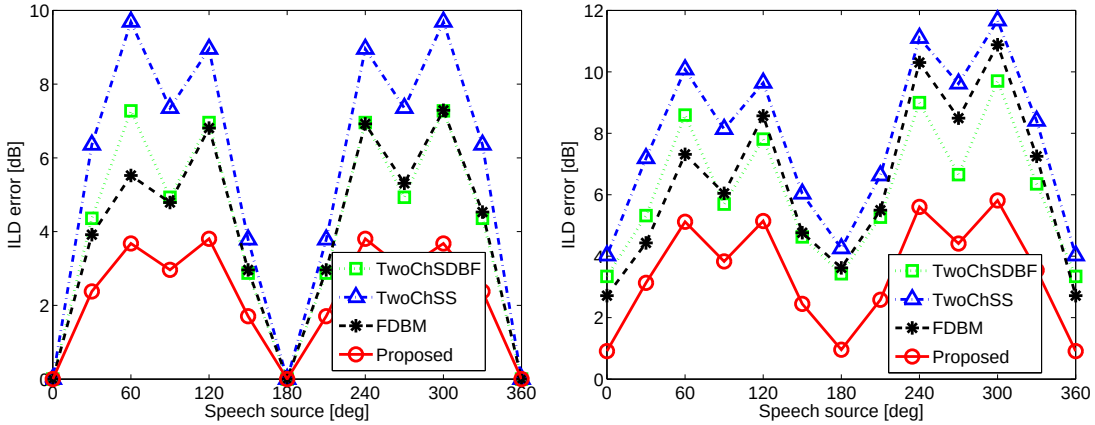
(a) (b)

Figure 4: MOS improvements (ΔMOS) of the studied algorithms at the left ear (a) and the right ear (b) in different acoustical conditions. The acoustical spatial configurations are presented in Table 1.



(a) (b)

Figure 5: The ITD errors (ΔE_{ITD}) in one-noise-source conditions ($S_{0:30:360}N_0$) (a) and three-noise-source conditions $S_{0:30:360}N_{90,180,270}$ (b).



(a) (b)

Figure 6: The ILD errors (ΔE_{ILD}) in one-noise-source conditions $S_{0:30:360}N_0$ (a) and three-noise-source conditions $S_{0:30:360}N_{90,180,270}$ (b).

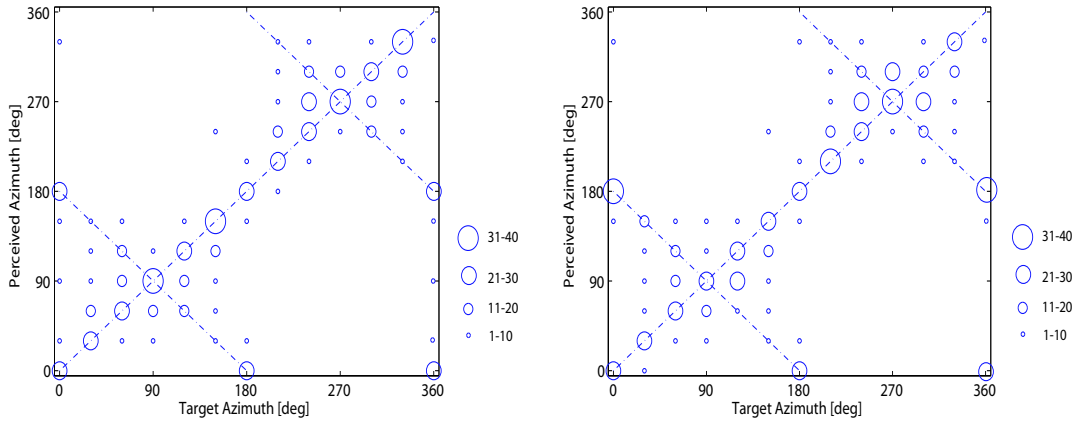


Figure 7: Results of subjective sound localization tests in one-noise-source conditions ($S_{0:30:360}N_0$) (a) and three-noise-source conditions ($S_{0:30:360}N_{90,180,270}$) (b).