11

Preservation of Localization Cues in BSS-Based Noise Reduction: Application in Binaural Hearing Aids

Jorge I. Marin-Hurtado[1] and David V. Anderson[2]
[1]*Universidad del Quindio, Department of Electronics Engineering, Armenia, Q.*
[2]*Georgia Institute of Technology, School of Electrical and Computer Engineering, Atlanta, GA*
[1]*Colombia*
[2]*USA*

1. Introduction

For speech applications, blind source separation provides an efficient strategy to enhance the target signal and to reduce the background noise in a noisy environment. Most ICA-based blind source separation (BSS) algorithms are designed under the assumption that the target and interfering signals are spatially located. When the number of interfering signals is small, one of the BSS outputs is expected to provide an excellent estimation of the target signal. Hence, the overall algorithm behaves as an "ideal" noise-reduction algorithm. However, when the number of interfering signals increases, problem known as the cocktail party effect, or when the background noise is diffusive (i.e., non-point-source noise), this BSS output is no longer a good estimation of the target signal. (Takahashi et al., 2009) showed that in a two-output ICA-based BSS algorithm under these adverse environments, one BSS output includes a mixture of the target signal and residual noise related to the interfering signals, while the other output provides an accurate estimation of the background noise. This particular property validates the experimental results achieved by different post processing strategies to enhance the BSS output associated to the target signal (Noohi & Kahaei, 2010; Parikh et al., 2010; Parikh & Anderson, 2011; Park et al., 2006). These methods are based on Wiener filtering (Kocinski, 2008; Noohi & Kahaei, 2010; Park et al., 2006), spectral subtraction (Kocinski, 2008), least-square (LS) minimization (Parikh et al., 2010), and perceptual post processing (Parikh & Anderson, 2011). All these methods take advantage of a reliable background noise estimator obtained at one of the BSS outputs.

The above BSS-based noise-reduction methods provide a single output, which means that the direction of arrival of the target signal (also known as binaural cue or localization cue) is lost in the enhanced signal. There are some applications such as the new generation of hearing aids, called binaural hearing aids, which demands noise-reduction algorithms that preserve localization cues. These binaural hearing aids are targeted for hearing-impaired people who suffer from hearing losses at both ears. A binaural hearing aid consists of two hearing devices, one per each ear, and a wireless link to exchange information between both hearing devices.

This wireless link can be used to synchronize the processing performed by both hearing aids or to exchange the signals received at each side. The latter allows the use of multi-microphone noise-reduction algorithms such as BSS-based noise reduction algorithms. The perceptual advantages of a binaural processing over independent non-synchronized hearing aids have been extensively documented by (Moore, 2007; Smith et al., 2008; Van den Bogaert et al., 2006). These perceptual studies showed subject preference for those algorithms that preserve the direction of arrival (localization cues) of the target and interfering signals. Hence, this chapter addresses the problem about the preservation of the localization cues in noise-reduction algorithms based on BSS, whose main target application is a binaural hearing aid.

This chapter includes an overview of the state-of-the-art BSS-based noise-reduction algorithms that preserve localization cues. This overview describes in detail five BSS algorithms to recover the localization cues: BSS constrained optimization (Aichner et al., 2007; Takatani et al., 2005), spatial-placement filter (Wehr et al., 2008; 2006), post processing based on adaptive filters (Aichner et al., 2007), post processing based on a Wiener filter (Reindl et al., 2010), and perceptually-inspired post processing (Marin-Hurtado et al., 2011; 2012). This chapter also discusses the advantages and limitations of each method, and presents the results of a comparative study conducted under different kinds of simple and adverse scenarios: multi-talker scenario, diffusive noise, babble noise. Performance of these algorithms is evaluated in terms of signal-to-noise ratio (SNR) improvement, subjective sound quality, and computational cost. The comparative study concludes that the perceptually-inspired post processing outperforms the adaptive-filter-based and the Wiener-filter-based post processing in terms of SNR improvement, noise reduction, and computational cost. Therefore, the perceptually-inspired post processing is outlined as a good candidate for the implementation of a binaural hearing aid. A discussion about the proposed future work and improvements in the proposed methods are also addressed at the end of this chapter.

2. The problem of preservation of localization cues in blind source separation

This section presents a general overview of a blind source separation (BSS) process, and its problem with respect to the spatial placement of the separated sources in the output of the BSS algorithm.

Suppose a BSS system with P sensors. In the frequency domain, a source signal $s_1(\omega)$ is perceived at the sensor array as

$$\boldsymbol{x}(\omega) = \begin{bmatrix} x_1(\omega) \\ \vdots \\ x_P(\omega) \end{bmatrix} = \boldsymbol{h}_1(\omega) s_1(\omega) \tag{1}$$

where $x_p(\omega)$, $p = 1, ..., P$, are the signals at each sensor, and $\boldsymbol{h}_1(\omega)$ is a vector that describes the propagation from the point source to each sensor. In particular for a hearing aid with one microphone per hearing device, i.e., $P = 2$, this vector is called the head-related transfer function (HRTF). In a binaural system, the preservation of these HRTFs is critical since they provide information to the human auditory system about the direction of arrival of the target signals.

When Q sources are present in the environment, the input vector $\boldsymbol{x}(\omega)$ at the sensor array is given by

$$\boldsymbol{x}(\omega) = \sum_{q=1}^{Q} \boldsymbol{h}_q(\omega) s_q(\omega) = \begin{bmatrix} h_{11}(\omega) & \cdots & h_{1Q}(\omega) \\ \vdots & \ddots & \vdots \\ h_{P1}(\omega) & \cdots & h_{PQ}(\omega) \end{bmatrix} \begin{bmatrix} s_{1(\omega)} \\ \vdots \\ s_Q(\omega) \end{bmatrix} = \boldsymbol{H}(\omega)\boldsymbol{s}(\omega), \tag{2}$$

where $\boldsymbol{H}(\omega) = [\boldsymbol{h}_1(\omega) \cdots \boldsymbol{h}_Q(\omega)]$, is called the mixing matrix, and the vector $\boldsymbol{s}(\omega)$ holds the frequency components of each source signal. For BSS-based noise-reduction applications, the source s_1 is typically assigned to the target signal, and the sources s_q, $q = 2, \ldots, Q$, are related to the interfering signals.

The purpose of any blind source separation algorithm is to recover the source signals $s_q(\omega)$ from the mixture $\boldsymbol{x}(\omega)$ by means of a linear operation denoted by the unmixing matrix $\boldsymbol{W}(\omega)$,

$$\boldsymbol{y}(\omega) = \begin{bmatrix} y_{1(\omega)} \\ \vdots \\ y_P(\omega) \end{bmatrix} = \begin{bmatrix} w_{11}(\omega) & \cdots & w_{1Q}(\omega) \\ \vdots & \ddots & \vdots \\ w_{P1}(\omega) & \cdots & w_{PQ}(\omega) \end{bmatrix} \begin{bmatrix} x_{1(\omega)} \\ \vdots \\ x_P(\omega) \end{bmatrix} = \boldsymbol{W}(\omega)\boldsymbol{x}(\omega), \tag{3}$$

where the elements of the matrix $\boldsymbol{W}(\omega)$ denote FIR filters designed to separate the source signals (Fig. 1). These filter weights are designed by an optimization process, where the minimization of the mutual information between the source signals is one of the most successful methods to derive these filter weights (Haykin, 2000). This chapter does not include a detailed description about the methods to estimate the unmixing matrix $\boldsymbol{W}$, except those to recover the localization cues in the BSS filter (Section 3.1).

The whole process can be described by

$$\boldsymbol{y}(\omega) = \boldsymbol{W}(\omega)\boldsymbol{H}(\omega)\boldsymbol{s}(\omega) = \boldsymbol{C}(\omega)\boldsymbol{s}(\omega), \tag{4}$$

where $\boldsymbol{C}(\omega) = \boldsymbol{W}(\omega)\boldsymbol{H}(\omega)$. When the number of the sources and sensors is identical, i.e., $P = Q$, the problem is well-posed, and the matrix $\boldsymbol{C}(\omega)$ becomes diagonal. In this case, $\boldsymbol{y}(\omega) \approx \boldsymbol{s}(\omega)$ or equivalently $y_p(\omega) = \hat{s}_p(\omega)$, $p = 1, \ldots, P$, and $\hat{s}_p(\omega)$ is an estimate of the source signal. Hence, the localization cues of each source signal are lost after the blind source separation. For example, if a binaural hearing aid with one microphone per hearing device, i.e., $P = 2$, is used to cancel out the interfering signal in an environment with one target and one interfering signal, i.e., $Q = 2$, the BSS outputs are expected to be $y_1(\omega) = \hat{s}_1(\omega)$ and $y_2(\omega) = \hat{s}_2(\omega)$. Then, the output $y_1(\omega)$ holds an estimate of the target signal. If the signal $y_1(\omega)$ is applied simultaneously to the left and the right ear, the signal is heard coming always from the front. To avoid this issue, a spatial placement of the estimate $\hat{s}_1$ is required at the output of the entire process. This recovery of the localization cues is described by

$$\boldsymbol{z}(\omega) = \begin{bmatrix} z_1(\omega) \\ z_2(\omega) \end{bmatrix} = \boldsymbol{h}_1(\omega)\hat{s}_1(\omega) \tag{5}$$

where z_1 and z_2 are the signals to deliver to the left and right channel, respectively, and $\boldsymbol{h}_1$ denotes the HRTF for the target signal. The above process can be performed by different approaches. A first approach is to modify the derivation of the BSS filter weights, $\tilde{\boldsymbol{W}}$, such as the output of the BSS algorithm, $\boldsymbol{z}(\omega) = \tilde{\boldsymbol{W}}(\omega)\boldsymbol{H}(\omega)\boldsymbol{s}(\omega)$ is constrained to be $\boldsymbol{z}(\omega) \approx$

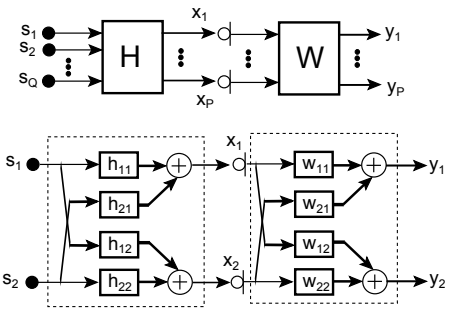

Fig. 1. General description of the blind source separation process for $P \neq Q$ (top); and two-sources and two-sensors $P = Q = 2$ (bottom).

$\boldsymbol{h}_1(\omega)s_1(\omega)$. These methods are discussed in the Section 3.1. Another approach, is to use a BSS post processing such as the output $y_1(\omega)$ is placed spatially by means of a filter $\boldsymbol{b}(\omega)$ such as

$$\boldsymbol{z}(\omega) = \boldsymbol{b}(\omega)y_1(\omega) \tag{6}$$

that ensures the fulfillment of the condition (5). This filter, called spatial-placement filter, is addressed in the Section 3.2. Another approach to recover the localization cues is to estimate a set of noise-suppression gains from the BSS outputs, and to apply these noise-suppression gains to the unprocessed signals. These methods are presented in the Sections 3.3 through 3.5, and are shown to provide more advantages than the BSS constrained optimization or the spatial-placement filter.

Up to this point the problem about recovery of the localization cues has been discussed for the case when $P = Q$ but in many practical applications, such as noise reduction, this condition cannot be met in adverse environments. In these environments, the number of interfering signals is larger than the number of sources, $Q > P$. This situation, called the undetermined case, leads to an ill-conditioned problem. Although the performance of the BSS algorithm is degraded for an undetermined case, the strategies to recover the localization cues in the undetermined case are exactly the same as described above. The main difference between both cases is regarding to the preservation of the localization cues for the interfering signals. These issues are described in detail in the next section.

3. BSS-based binaural noise-reduction algorithms

Most BSS algorithms are designed to support more than two sensors. As a general rule, increasing the number of sensors can separate more interfering sources but at expenses of increasing the computational complexity. In speech enhancement applications, for some adverse environments, the number of interfering signals is typically larger than the number of sources, $Q > P$, or even worse, the interfering signals are non-point noise sources, e.g., babble noise. Hence, increasing the number of sensors cannot improve significantly the quality of the source separation performed by the BSS algorithm. For this reason, a wide range of BSS-based speech enhancement algorithms are proposed for two-output BSS systems. Using two-output BSS algorithms provides additional advantages for some applications such as binaural hearing aids since the computational complexity and the wireless-link bandwidth can be reduced.

When two-output BSS algorithms are used in noise-reduction applications, the primary BSS output provides an estimate of the target signal, and the secondary BSS output, an estimate of the interfering signals. However, the estimate of the target signal does not provide information about the direction of arrival, and additional strategies are required to recover these localization cues. The approaches proposed in (Takatani et al., 2005) and (Aichner et al., 2007) employ a constrained optimization to derive the BSS filter weights. Unfortunately, these methods have shown a poor performance based on subjective tests (Aichner et al., 2007). More recent approaches use a BSS post processing stage to recover the localization cues and to enhance the target signal (Aichner et al., 2007; Marin-Hurtado et al., 2011; Reindl et al., 2010; Wehr et al., 2006). In these post-processing methods, the BSS outputs are used to compute noise-suppression gains that enhance the target signal. These post-processing methods have shown to be successful in the recovery of the localization cues and the reduction of the background noise, which is explained by the theoretical analysis conducted in (Takahashi et al., 2009) for two-output ICA-based BSS algorithms. In (Takahashi et al., 2009), authors showed that the estimate of the interfering signals is close to the true value whereas the estimate of the target signal includes a large amount of residual noise. Hence, when the estimate of the interfering signals is used in the post processing stage to compute the noise-suppression gains, the background noise can be significantly reduced.

In the BSS post-processing methods, depending on how these noise-suppression gains are applied to obtain the enhanced signal, it is possible to distinguish two groups. In the first group, these gains are applied to enhance the BSS output corresponding to the estimate of the target signal (Fig. 2a). In the second group, these gains are applied directly to the unprocessed signals (Fig. 2b). In BSS-based binaural speech enhancement applications, these noise-suppression gains are used not only to enhance the speech signal but also to recover the direction of arrival (or localization cues) of the speech signal. Although both groups of post processing are successful to recover the localization cues of the target signal, experimental and theoretical analysis show that the first group, in which BSS noise-suppression gains are applied to the BSS outputs, cannot recover the localization cues for the interfering signals (Aichner et al., 2007; Marin-Hurtado et al., 2012; Reindl et al., 2010; Wehr et al., 2008). In this case, the interfering signals are usually mapped to the direction of arrival of the target signal. This effect is not a desirable feature for binaural hearing aids, in which the displacement of the localization cues is identified as annoying through perceptual experiments (Moore, 2007; Smith et al., 2008; Sockalingam et al., 2009). On the other hand, the BSS post-processing methods that apply the noise-reduction gains to the unprocessed signals are shown to be successful in the recovery of the localization cues for both target and interfering signals simultaneously (Marin-Hurtado et al., 2012; Reindl et al., 2010).

3.1 BSS constrained optimization

As mentioned in the Section 2, localization cues can be recovered by using a constrained optimization in the derivation of the BSS filter weights, $\tilde{\boldsymbol{W}}$, such as the BSS output $\boldsymbol{z}(\omega) = \tilde{\boldsymbol{W}}(\omega)\boldsymbol{H}(\omega)\boldsymbol{s}(\omega)$ is constrained to be $\boldsymbol{z}(\omega) \approx \boldsymbol{h}_1(\omega)s_1(\omega)$, where s_1 and $\boldsymbol{h}_1$ are the target signal and its HRTF.

In (Takatani et al., 2005), authors proposed a BSS algorithm using the structure shown in Fig. 3, which uses a cost function that involves two terms,

$$\mathcal{J}(n) = \mathcal{J}_{\boldsymbol{y}}(n) + \beta\mathcal{J}_{\tilde{\boldsymbol{y}}}(n). \tag{7}$$

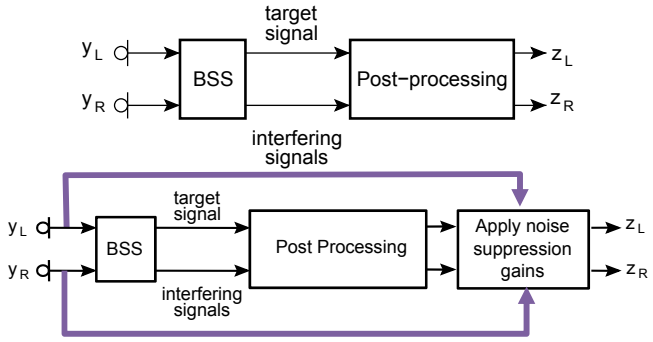

Fig. 2. BSS post processing to recover the localization cues: Post processing that enhances the BSS output (top), and post processing that enhances the unprocessed signals (bottom).

The first term, $\mathcal{J}_{\boldsymbol{y}}(n)$, is related to the classical source separation algorithms by minimization of the mutual information between the output channels y_1 and y_2, $\boldsymbol{y}(n) = [y_1(n)\, y_2(n)]^T$, and the second term, $\mathcal{J}_{\tilde{\boldsymbol{y}}}(n)$, is the minimization of the mutual information between the combination of the channels, $\tilde{y}(n) = [\tilde{y}_1(n)\, \tilde{y}_2(n)]^T$,

$$\begin{aligned} \tilde{y}_1(n) &= x_1(n-l) - y_1(n) \\ \tilde{y}_2(n) &= x_2(n-l) - y_2(n), \end{aligned}$$

where l is a time delay to compensate the processing delay introduced by the unmixing filters $\boldsymbol{w}$, and the parameter β controls a trade-off between both cost functions. The cost functions $\mathcal{J}_{\boldsymbol{y}}(n)$ and $\mathcal{J}_{\tilde{\boldsymbol{y}}}(n)$ are based on the statistical independence measurement given by the Kullback-Leibler divergence (KLD) or relative entropy, (Takatani et al., 2005)

$$\mathcal{J}_{\boldsymbol{y}}(n) = \mathcal{E}\left\{ \log \frac{\hat{p}_{y,P}(\boldsymbol{y}(n))}{\prod_{q=1}^{P} \hat{p}_{y,1}(y_q(n))} \right\} \tag{8}$$

and

$$\mathcal{J}_{\tilde{\boldsymbol{y}}}(n) = \mathcal{E}\left\{ \log \frac{\hat{p}_{y,P}(\tilde{\boldsymbol{y}}(n))}{\prod_{q=1}^{P} \hat{p}_{y,1}(\tilde{y}_q(n))} \right\} \tag{9}$$

where $\hat{p}_{y,P}(.)$ is the estimate of the P-dimensional joint probability density function (pdf) of all channels, $\hat{p}_{y,1}(.)$ is the estimate of the the uni-variate pdfs, and $\mathcal{E}\{.\}$ is the expected value.

A disadvantage of the Takatani *et al.*'s method is the huge computational cost and the slow convergence. An alternative solution proposed by (Aichner et al., 2007) replaces the minimization of the mutual information of the combined channels, $\tilde{\boldsymbol{y}}$, by a minimization of the minimum mean-square error (MMSE) of the localization cues,

$$\mathcal{J}(n) = \mathcal{J}_{\boldsymbol{y}}(n) + \gamma \mathcal{E} \left\| \boldsymbol{x}(n-l) - \boldsymbol{y}(n) \right\|^2, \tag{10}$$

where $\mathcal{J}_{\boldsymbol{y}}(n)$ is given by (8), γ is a trade-off parameter, and l is a time delay to compensate the processing delay introduced by the BSS algorithm (Fig. 4). The rationale behind the above method is that localization cues of the target signal in BSS inputs, $\boldsymbol{x}(n)$, must be kept in the BSS outputs, $\boldsymbol{y}(n)$, which is equivalent to minimize the MMSE between the input and output.

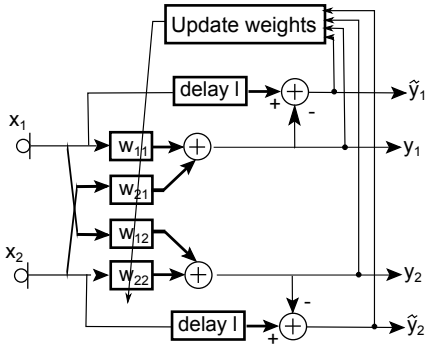

Fig. 3. Block diagram of the BSS constrained optimization to recover the localization cues proposed by (Takatani et al., 2005).

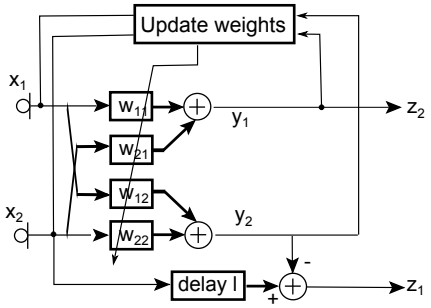

Fig. 4. Block diagram of the BSS constrained optimization to recover the localization cues proposed by (Aichner et al., 2007).

Although the subjective test conducted in (Aichner et al., 2007) showed that both methods can preserve the localization cues of the target signal, both methods cannot preserve the localization cues of the suppressed interfering signals, and the interfering signals are heard strongly distorted. In terms of noise reduction, the BSS constrained optimization method proposed in (Aichner et al., 2007) provides better performance than (Takatani et al., 2005).

3.2 Post processing based on spatial-placement filter

The main disadvantage of the BSS constrained optimization algorithms is their high computational cost. This issue can be solved by the spatial-placement filter introduced in (6), Section 2. A block diagram of the spatial-placement filter is shown in Fig. 5. The purpose of this filter is to recover the localization cues that are lost in the BSS output related to the target signal. If the BSS output holding the estimate of the target signal is $y_1(\omega)$, the spatial-placement filter, $\boldsymbol{b}(\omega)$, $\boldsymbol{z}(\omega) = \boldsymbol{b}(\omega)y_1(\omega)$ must satisfy (5), i.e., in the ideal case,

$$\boldsymbol{b}(\omega)y_1(\omega) = \boldsymbol{h}_1(\omega)s_1(\omega). \tag{11}$$

According to (2), the HRTF $\boldsymbol{h}_1(\omega)$ corresponds to the first column of the mixing matrix $\boldsymbol{H}(\omega)$,

$$\boldsymbol{h}_1(\omega) = \boldsymbol{H}(\omega)\boldsymbol{e}_1 \tag{12}$$

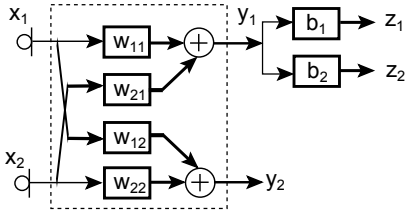

Fig. 5. Block diagram of the spatial-placement filter to recover the localization cues.

with $\boldsymbol{e}_1 = [1\,0\,...\,0]^T$. From (4),

$$\boldsymbol{H}(\omega) = \boldsymbol{W}^{-1}(\omega)\boldsymbol{C}(\omega), \tag{13}$$

Thus, replacing (12) and (13) in (11),

$$\boldsymbol{b}(\omega)y_1(\omega) = \boldsymbol{W}^{-1}(\omega)\left[\boldsymbol{C}(\omega)\boldsymbol{e}_1 s_1(\omega)\right] \tag{14}$$

where the term in brackets $\boldsymbol{C}(\omega)\boldsymbol{e}_1 s_1(\omega) = \boldsymbol{W}(\omega)\boldsymbol{H}(\omega)\boldsymbol{e}_1 s_1(\omega) = \boldsymbol{W}(\omega)\boldsymbol{h}_1(\omega)s_1(\omega)$ is the output of the BSS algorithm when only the target signal is present in the environment. In other terms, the term in brackets becomes $\boldsymbol{e}_1 y_1(\omega)$. Thus,

$$\boldsymbol{b}(\omega) = \boldsymbol{W}^{-1}(\omega)\boldsymbol{e}_1 \tag{15}$$

or in other words, the coefficients of the spatial-placement filter correspond to the first column of the inverse matrix $\boldsymbol{W}^{-1}$.

A practical implementation of (15) requires the regularization of the inverse matrix to avoid an unstable algorithm. However, even using this regularization, the method in (15) is impractical for the recovery of the localization cues (Wehr et al., 2006). For example, suppose an environment with two sources, one target signal, s_1, and one interfering signal, s_2. In this environment, the signals perceived in the sensor array are given by

$$\boldsymbol{x} = \boldsymbol{h}_1 s_1 + \boldsymbol{h}_2 s_2 = \begin{bmatrix} h_{11} \\ h_{21} \end{bmatrix} s_1 + \begin{bmatrix} h_{12} \\ h_{22} \end{bmatrix} s_2 \tag{16}$$

Hence, in an ideal binaural noise-reduction system, the spatial-placement filter is expected to provide an output with structure similar as (16) but scaling down the term related to the interfering signal.

If a two-output BSS algorithm is used to cancel out the interfering signal, the output of the spatial-placement filter,

$$\boldsymbol{z}(\omega) = \boldsymbol{W}^{-1}(\omega)\boldsymbol{e}_1 y_1(\omega) = \boldsymbol{W}^{-1}(\omega)\boldsymbol{e}_1 \sum_{j=1}^{P} c_{1j}s_j\ , \tag{17}$$

is described in terms of the matrix elements c_{ij} and h_{ij} as

$$\boldsymbol{z} = \begin{bmatrix} h_{11} \\ h_{21} \end{bmatrix} s_1 - \frac{c_{21}}{c_{22}} \begin{bmatrix} h_{12} \\ h_{22} \end{bmatrix} s_1 + \frac{c_{12}}{c_{11}} \begin{bmatrix} h_{11} \\ h_{21} \end{bmatrix} s_2 \tag{18}$$

where the above derivation used the facts that $\boldsymbol{W}^{-1}(\omega) = \boldsymbol{H}(\omega)\boldsymbol{C}^{-1}(\omega)$, and $\boldsymbol{C}$ becomes a diagonal matrix in the determined case, i.e., $\|c_{11}c_{22}\|^2 \gg \|c_{12}c_{21}\|^2$. In the above equations, the variable ω is omitted for mathematical convenience. In (18) is clear that the target signal, s_1, is mapped to the desired direction of arrival, $\boldsymbol{h}_1 s_1 = [h_{11}\, h_{21}]^H s_1$. On the other hand, the interfering signal, s_2, is scaled by a factor c_{12}/c_{11} but it is also mapped to the direction of arrival of the target signal, which suggests that the localization cues for s_2 are not preserved. Another critical problem of this spatial-placement filter arises from the second term in (18). This term suggests that the target signal, s_1, is also mapped to the direction of arrival of the interfering signal and scaled by a factor c_{21}/c_{22}.

To avoid the regularization of the inverse matrix $\boldsymbol{W}^{-1}$ and the mapping of the target signal into the direction of arrival of the interfering signal, (Wehr et al., 2008; 2006) proposed to use the adjoint of the mixing matrix, $\boldsymbol{H}$, as unmixing matrix, i.e., $\boldsymbol{W}(\omega) = adj\{\boldsymbol{H}(\omega)\}$. Under this assumption, the spatial-placement filter that satisfies (11) is given by

$$\boldsymbol{b}(\omega) = adj\{\boldsymbol{W}(\omega)\}\,\boldsymbol{e}_1 \tag{19}$$

Then the output of the spatial-placement filter is given by

$$\boldsymbol{z}(\omega) = adj\{\boldsymbol{W}(\omega)\}\,\boldsymbol{e}_1 \sum_{j=1}^{P} c_{1j} s_j \tag{20}$$

Again, for an environment with one target and one interfering signal, the output of the spatial-placement filter of a two-output BSS algorithm is given by (Wehr et al., 2006)

$$\boldsymbol{z} = det\{\boldsymbol{W}(\omega)\}\left(\begin{bmatrix} h_{11} \\ h_{21} \end{bmatrix} s_1 + \frac{c_{12}}{c_{11}} \begin{bmatrix} h_{11} \\ h_{21} \end{bmatrix} s_2\right). \tag{21}$$

This equation shows that localization cues of the target signal, s_1, can be recovered correctly. However, the localization cues of the interfering signal are lost since the interfering signal is mapped to the direction of arrival of the target signal. The effect of this displacement in the localization cues for the interfering signal was evaluated in (Aichner et al., 2007) by a subjective test. Results showed that the post processing based on spatial-placement filter can be outperformed by a post processing based on adaptive filter, which is discussed in the next section.

3.3 Post processing based on adaptive filter

Up to this point the approaches discussed to recover the localization cues, BSS constrained optimization and BSS post processing using spatial-placement filter, fail to recover the localization cues of the interfering signals even under the determined case, i.e., when the number of source signals and sensors is the same, $P = Q$. In these methods, the localization cues of the interfering signals are usually mapped to the direction of arrival of the target signal.

To avoid the displacement of the localization cues for the interfering signals, different authors have reported the use of noise-suppression gains applied to the unprocessed signals rather than apply noise-suppression gains to the BSS outputs as in the spatial-placement filter. The first approach proposed to recover efficiently the localization cues was reported by (Aichner et al., 2007), which uses adaptive filters to cancel out the background noise. A block diagram of the method proposed in (Aichner et al., 2007) is shown in Fig. 6. In this approach, a BSS

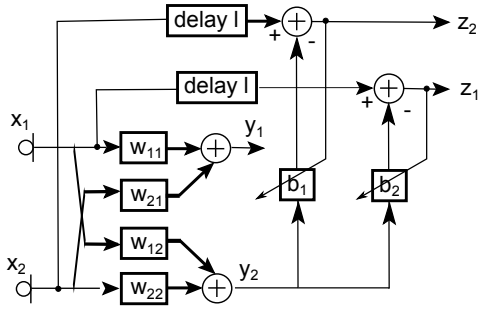

Fig. 6. BSS post processing based on adaptive filters. In this figure y_2 provides an estimate of the interfering signals $\hat{u}(n)$.

algorithm provides an estimate of the interfering signals, $\hat{u}(n)$, and this estimate is used as input for two adaptive filters, one for each side. The desired inputs for these adaptive filters are the unprocessed signals at the left and right channel. Then, the error signals provide enhanced signals in which the localization cues can be preserved.

This post processing can be used together any BSS algorithm. The original description of this algorithm uses the BSS algorithm described in (Aichner et al., 2006). On the other hand, the adaptive filters are designed in the DFT domain to minimize the time-averaged error (Aichner et al., 2007):

$$\mathcal{J}_{AF}(n) = (1-\lambda)\sum_{i=0}^{n}\lambda^{n-i}\sum_{k=0}^{R-1}\left|z_p(k,i)\right|^2 \tag{22}$$

where $z_p(k,i)$, $p \in \{1,2\}$, represents the DFT of the output of the algorithm at the frequency bin k and time index i; $0 < \lambda < 1$ is a forgetting factor; and R is the DFT length. The filter coefficients derived from (22) are given by

$$b_p(k,n) = \frac{r_{ux}(k,n)}{r_{uu}(k,n)} \tag{23}$$

where

$$\begin{aligned} r_{ux}(k,n) &= \lambda r_{ux}(k,n-1) + x_p(k,n)\,u(k,n) \\ r_{uu}(k,n) &= \lambda r_{uu}(k,n-1) + |u(k,n)|^2 \ ; \end{aligned}$$

$x_p(k,n)$ is the DFT of the input signal at the frequency bin k, time index n, and microphone p; and $u(k,n)$ is the DFT of the BSS output related to the interfering signals.

3.3.1 Limitations

In (Aichner et al., 2007), authors compared the BSS constrained optimizations given in (7) and (10), the spatial-placement filter given in (20), and the post processing based on adaptive filters given in (23), concluding that the post processing based on adaptive filters outperforms the other methods and preserves efficiently the localization cues.

The experimental results of the Aichner's study were conducted only for environments with two sources. Further research identified some problems in the BSS post processing based

on adaptive filters. In (Reindl et al., 2010), a theoretical analysis of the adaptive-filter-based post processing shows that the noise reduction can be performed efficiently only under the determined case, i.e., when $P \geq Q$. In the undetermined case, $P < Q$, the noise reduction is possible only if the interfering signals are located at the same position.

To show the above statements lets consider a two-input BSS algorithm. In this algorithm we assume that the BSS output y_2 holds the estimate of the interfering signals, $\hat{u}(\omega) = y_2(\omega)$. This estimate in the frequency domain is given by

$$\hat{u}(\omega) = w_{11}(\omega)x_1(\omega) + w_{21}(\omega)x_2(\omega) = \sum_{p=1}^{2} w_{p1}(\omega)x_p(\omega). \tag{24}$$

In the general case, $x_p(\omega)$ is described by (2),

$$x_p(\omega) = \boldsymbol{e}_p^T \sum_{q=1}^{Q} \boldsymbol{h}_q(\omega)s_q(\omega) = \sum_{q=1}^{Q} h_{pq}(\omega)s_q(\omega). \tag{25}$$

Independent on the algorithm selected for the BSS algorithm, the target signal, s_1, can be assumed to be perfectly canceled out in $\hat{u}(\omega)$, which is expressed through

$$\hat{u}(\omega) = \sum_{p=1}^{2} w_{p1}(\omega) \sum_{q=2}^{Q} h_{pq}(\omega)s_q(\omega) \tag{26}$$

The output of the adaptive filters can be obtained by means of

$$z_p(\omega) = x_p(\omega) - b_p(\omega)\hat{u}(\omega) \quad p \in \{1, 2\}. \tag{27}$$

Thus, replacing (25) and (26) in (27),

$$z_p(\omega) = h_{1p}(\omega)s_1(\omega) + \sum_{q=2}^{Q} \left[h_{qp}(\omega) - b_p(\omega)c_q(\omega)\right] s_q(\omega) \tag{28}$$

where

$$c_q(\omega) = w_{11}(\omega)h_{q1}(\omega) + w_{21}(\omega)h_{q2}(\omega). \tag{29}$$

From (28), to cancel out all interfering point sources, the frequency response of the adaptive filters must satisfy the condition

$$\sum_{q=2}^{Q} \left[h_{qp}(\omega) - b_p(\omega)c_q(\omega)\right] s_q(\omega) = 0 \tag{30}$$

In the determined case, $P = Q = 2$, the above equation can be satisfied if $b_p(\omega) = \frac{h_{2p}(\omega)}{c_2(\omega)}$. In the non-determined case, $Q > P$, it is necessary to satisfy the following simultaneous conditions,

$$b_p(\omega) = \frac{h_{2p}(\omega)}{c_2(\omega)} \cap b_p(\omega) = \frac{h_{3p}(\omega)}{c_3(\omega)} \cap \cdots \cap b_p(\omega) = \frac{h_{Qp}(\omega)}{c_Q(\omega)} \tag{31}$$

or equivalently,

$$\frac{h_{2p}(\omega)}{c_2(\omega)} = \frac{h_{3p}(\omega)}{c_3(\omega)} = \cdots = \frac{h_{Qp}(\omega)}{c_Q(\omega)}$$

For the particular case of two interfering sources, $Q = 3$, and two microphones, $P = 2$,

$$\frac{h_{2p}(\omega)}{c_2(\omega)} = \frac{h_{3p}(\omega)}{c_3(\omega)}$$

which leads to

$$w_{21}(\omega)\left[h_{21}(\omega)h_{32}(\omega) - h_{22}(\omega)h_{31}(\omega)\right] = 0$$
$$w_{11}(\omega)\left[h_{22}(\omega)h_{31}(\omega) - h_{21}(\omega)h_{32}(\omega)\right] = 0$$

Avoiding the trivial solution, these equations are true if $h_{21}(\omega)h_{32}(\omega) - h_{22}(\omega)h_{31}(\omega) = 0$, i.e., only if the interfering sources are located at the same position since $h_{21}(\omega) = h_{31}(\omega)$ and $h_{32}(\omega) = h_{22}(\omega)$. Hence, the performance of this post-processing method is fair in multiple-source environments such as babble noise.

Furthermore, a subjective evaluation in (Marin-Hurtado et al., 2011) showed that the adaptive-filter-based post processing cannot preserve the localization cues in the undetermined case. In this case, the interfering signals are mapped to the direction of arrival of the target signal. These experimental findings are explained by a mathematical derivation in (Marin-Hurtado et al., 2012), which is based on an analysis of the interaural transfer function (ITF). The magnitude of the ITF is called interaural level differences (ILD), and its phase is called interaural time differences (ITD). To preserve the localization cues, any post-processing method should ensure an output ITF similar to the input ITF for all frequencies, i.e., $ITF^{in}(\omega) = ITF^{out}(\omega)\ \forall\omega$. These ITFs are defined by the ratios

$$ITF^{in}(\omega) = \frac{x_1(\omega)}{x_2(\omega)}\ ;\ ITF^{out}(\omega) = \frac{z_1(\omega)}{z_2(\omega)} \tag{32}$$

In the post processing based on adaptive filters, the input and output ITF for every interfering signal are defined as

$$ITF_q^{in}(\omega) \triangleq \frac{h_{q1}(\omega)}{h_{q2}(\omega)}\ ;\ ITF_q^{out}(\omega) \triangleq \frac{y_{q1}(\omega)}{y_{q2}(\omega)} \tag{33}$$

where

$$y_{qp}(\omega) = \left[h_{qp}(\omega) - b_p(\omega)c_q(\omega)\right] s_q(\omega).$$

Thus,

$$ITF_q^{out}(\omega) = ITF_q^{in}(\omega) + D_q(\omega)$$

where $q = 2, ..., Q$ and

$$D_q(\omega) = \frac{\left[b_2(\omega)h_{q1}(\omega) - b_1(\omega)h_{q2}(\omega)\right] c_q(\omega)}{\left[h_{q2}(\omega) - b_2(\omega)c_q(\omega)\right] h_{q2}(\omega)}$$

is the ITF displacement. In other words, the perceived direction of arrival for each interfering signal is shifted from its original position. In the determined case, the conditions given by (31) are satisfied, which leads to an ITF displacement $D_q(\omega) = 0$. On the other hand, an ITF displacement $D_q(\omega) \neq 0$ is obtained in the undetermined case since the conditions (31) are not met.

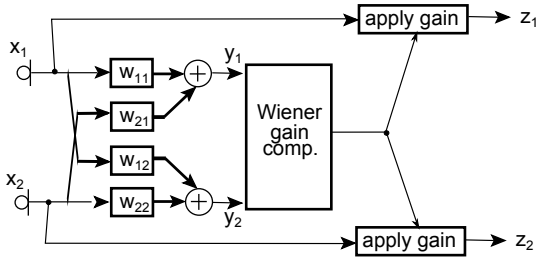

Fig. 7. Post processing based on Wiener filter.

3.4 Post processing based on Wiener filter

The methods described in the previous sections cannot preserve the localization cues for both target and interfering signals simultaneously in the undetermined case. In most of the cases, the interfering signals are mapped to the direction of arrival of the target signal. For other algorithms, such as the post processing based on adaptive filters, the localization cues can only be preserved under certain conditions in which the number of source signals is equal or lower than the number of sensors (determined case). From the perceptual viewpoint, these methods are impractical for binaural hearing aids since the displacement of the localization cues has been identified as annoying for hearing-impaired subjects.

In (Reindl et al., 2010), authors proposed an alternative post-processing stage based on Wiener filter to recover the localization cues. In this method, the BSS outputs are used to compute the Wiener filter gains, and these gains are applied simultaneously to the unprocessed signals (Fig. 7). This method is based on the fact that an ICA-based BSS algorithm provides a good estimate for the interfering signals, i.e., the BSS algorithm provides a good noise estimator. Since the Wiener filter gains are applied symmetrically to both sides, this method is ensured to preserve the localization cues for both target and interfering signals simultaneously.

The Wiener filter gains are computed by (Reindl et al., 2010)

$$g_{Reindl}(\omega) = \max\left\{1 - \alpha_\omega \frac{S_{\hat{n}\hat{n}}(\omega)}{S_{v_1v_1}(\omega)S_{v_2v_2}(\omega)}, 1\right\} \tag{34}$$

where $S_{\hat{n}\hat{n}}(\omega)$, $S_{v_1v_1}(\omega)$, and $S_{v_2v_2}(\omega)$ are the power spectral densities (PSD) of the estimate of the interfering signals (26), and the outputs of the intermediate unmixing filters $v_1(\omega) = w_{11}(\omega)x_1(\omega)$ and $v_2(\omega) = w_{21}(\omega)x_2(\omega)$. If the BSS output that holds the noise estimate $\hat{n}(\omega)$ is $y_1(\omega)$, the signals $v_1(\omega)$ and $v_2(\omega)$ take the forms $v_1(\omega) = w_{12}(\omega)x_1(\omega)$ and $v_2(\omega) = w_{22}(\omega)x_2(\omega)$. These PSDs can be updated by means of a first order estimator,

$$\begin{aligned}
S_{\hat{n}\hat{n}}(\omega, n) &= \lambda S_{\hat{n}\hat{n}}(\omega, n-1) + (1-\lambda)\,|\hat{n}(\omega, n)|^2 \\
S_{v_1v_1}(\omega, n) &= \lambda S_{v_1v_1}(\omega, n-1) + (1-\lambda)\,|w_{11}(\omega)x_1(\omega)|^2 \\
S_{v_2v_2}(\omega, n) &= \lambda S_{v_2v_2}(\omega, n-1) + (1-\lambda)\,|w_{21}(\omega)x_2(\omega)|^2
\end{aligned}$$

where λ is a time constant to smooth the estimator, and α_ω is a frequency-dependent trade-off parameter to control the roll-off of the noise reduction. Finally, the enhanced outputs are obtained by

$$\begin{aligned}
z_1(\omega) &= g_{Reindl}(\omega)x_1(\omega) \\
z_2(\omega) &= g_{Reindl}(\omega)x_2(\omega)
\end{aligned}$$

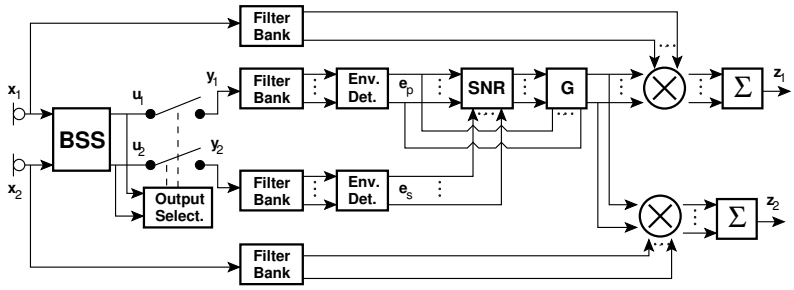

Fig. 8. Perceptually-inspired post processing to preserve the localization cues.

Experimental results in (Reindl et al., 2010) and (Marin-Hurtado et al., 2012) showed that this method can preserve the localization cues for both target and interfering signals simultaneously; however, the performance of this method is slightly below the performance of the post processing based on adaptive filters (Marin-Hurtado et al., 2012).

3.5 Perceptually-inspired post processing

In the previous sections, different BSS post-processing methods were discussed to recover the localization cues. All the above methods can preserve the localization cues of the target signal efficiently. However, only the BSS post-processing method based on Wiener filter can preserve the localization cues for both target and interfering signals simultaneously. This section discusses an alternative BSS post-processing method that preserves both localization cues. In this case, a perceptually-inspired post processing (BSS-PP) is used to compute a set of time-domain gains from the BSS outputs, and these gains are applied to the unprocessed signals (Fig. 8) (Marin-Hurtado et al., 2011; 2012). The BSS post processing used in (Marin-Hurtado et al., 2011; 2012) is an adaptation of the method in (Parikh & Anderson, 2011). This post processing is selected since it outperforms other BSS post processing for monaural speech enhancement applications. This post processing is modified so that it can be used for a binaural hearing aid (Marin-Hurtado et al., 2012):

1. To preserve the localization cues, the gains obtained by the BSS and perceptual post-processing algorithm described in (Parikh & Anderson, 2011) are applied to the unprocessed signals received at each side (Figure 8).
2. To achieve low processing delay, the system is implemented assuming real-time operating constraints, with the envelopes (e_p and e_s), SNR estimates, and gain parameters updated in the frame-by-frame basis, while the gains and outputs are computed in the sample-by-sample basis. In (Parikh & Anderson, 2011), gains are computed assuming an entire knowledge of the signal.
3. To minimize artifacts and to achieve more quality outputs, it is necessary to hold a long-term history for the maximum values of the primary envelope (e_p). Different tests show that the length of this memory should be at least one second.
4. To estimate the SNR, first-order estimators of the signal and noise PSD are used, and the SNR is computed as the ratio of these PSDs.

This perceptually-inspired BSS post processing is shown in Figure 8. Signals received at the left, x_1, and right, x_2, microphones are passed through a BSS algorithm to get u_1 and u_2. An

output selection algorithm identifies which BSS output contains the separated target signal (y_1), or primary channel, and the separated interfering signal (y_2), or secondary channel. These outputs, y_1 and y_2, are analyzed using an auditory filter bank, and then, the envelope in each sub-band is extracted. These envelopes are used to estimate the SNR and to compute the noise-suppression gains. The SNR and gains are computed separately for each sub-band. These noise-suppression gains expand the dynamic range of each sub-band by lowering the noise floor. These gains are finally applied simultaneously to the unprocessed signals by time-domain multiplication, and the outputs from each sub-band are summed together to produce the enhanced signals for the left and right ear.

To reduce computational complexity and processing delay in the BSS stage, an info-max BSS algorithm that uses adaptive filters to minimize the mutual information of the system outputs is used. This algorithm is described by the following set of equations (Marin-Hurtado et al., 2012):

$$u_1(n+1) = x_1(n) + \boldsymbol{w}_{12}^T(n)\boldsymbol{u}_2(n) \tag{35}$$

$$u_2(n+1) = x_2(n) + \boldsymbol{w}_{21}^T(n)\boldsymbol{u}_1(n) \tag{36}$$

$$\boldsymbol{w}_{12}(n+1) = \boldsymbol{w}_{12}(n) - 2\mu \tanh(u_1(n+1))\boldsymbol{u}_2(n) \tag{37}$$

$$\boldsymbol{w}_{21}(n+1) = \boldsymbol{w}_{21}(n) - 2\mu \tanh(u_2(n+1))\boldsymbol{u}_1(n)\,, \tag{38}$$

where x_1 and x_2 are the signals received at the left and right microphones, $\boldsymbol{w}_{12}$ and $\boldsymbol{w}_{21}$ are vectors of length N_w describing the unmixing filter coefficients, and $\boldsymbol{u}_1(n)$ and $\boldsymbol{u}_2$ are vectors of length N_w whose elements are the previous outputs of the BSS algorithm, $\boldsymbol{u}_j(n) = [u_j(n)\, u_j(n-1) \cdots u_j(n-N_w+1)]^T$, $j = 1, 2$, and n is the time index. To determine which BSS output contains the target signal, the time-average energy of the envelopes of the signals u_1 and u_2 are compared, and then, the output with higher time-average energy is selected as primary channel y_1. This time-average energy is computed by

$$u_j^{env}(n) = \eta_{env} u_j^{env}(n-1) + (1-\eta_{env}) u_j^2(n) \tag{39}$$

where η_{env} is a time constant. This update takes place every N samples.

The outputs of the BSS algorithm, $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$, as well as the unprocessed input signals at the left and right microphones, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, are passed through a filter bank that resembles the auditory system. This filter bank was implemented using forth-order Butterworth filters. At 22 kHz sampling rate, each filter bank provides 24 sub-bands. At the output of the filter banks, the vectors $\boldsymbol{x}_j(l,k)$ and $\boldsymbol{y}_j(l,k)$ of length N, $j = 1, 2$, are obtained, where l corresponds to the frame index and k to the sub-band number. Although the signals x and y are obtained in the sample-by-sample basis, they are analyzed in non-overlapped frames of length N to compute the gain parameters as we will show next.

For each output $\boldsymbol{y}_j(l,k)$, the envelope is extracted using a full-wave rectifier followed by a low-pass filter. In particular, the primary envelope vector $\boldsymbol{e}_p(l,k)$ is extracted from $\boldsymbol{y}_1(l,k)$, and the secondary envelope vector $\boldsymbol{e}_s(l,k)$ from $\boldsymbol{y}_2(l,k)$. The low-pass filters are implemented using a first-order IIR filter whose cutoff frequency is selected to be a fraction of the corresponding bandwidth of the band (Parikh & Anderson, 2011). These cutoff frequencies are set to $1/5$, $1/8$ and $1/15$ of the bandwidth of low, medium and high-frequency bands, respectively. These fractions ensure that the envelope tracks the signal closely but at the same time does not change too rapidly to cause abrupt gain changes that introduce modulation.

The final outputs at the left, z_1, and the right, z_2, side are computed using the time-domain gains $\boldsymbol{g}_{l,k}$ produced by the perceptual post-processing stage:

$$\boldsymbol{z}_j(l) = \sum_k \boldsymbol{g}_{l,k} \circ \boldsymbol{x}_j(l,k) \tag{40}$$

where $\circ$ denotes the element-wise product. The vector form emphasizes that the gains are computed using parameters updated on a frame-by-frame basis. However, these outputs can be computed on a sample-by-sample, reducing the processing delay.

In (Parikh & Anderson, 2011), inspired by a perceptual modeling, these gains modify the envelope of each sub-band $e_k(t)$ such that $\hat{e}_k(t) = \beta e_k^{\alpha}(t)$. To provide noise reduction, the maximum envelope value is preserved (i.e., $\hat{e}_{k_{max}} = e_{k_{max}}$) while the minimum envelope value is lowered (i.e., $\hat{e}_{k_{min}} = K e_{k_{min}}$, where K is an expansion coefficient). Using the previous ideas, (Parikh & Anderson, 2011) developed a method to estimate α and β from the entire signal. To provide a realistic implementation, equations in (Parikh & Anderson, 2011) are modified to a vector form to state the update of α and β is the frame-by-frame basis every N samples (Marin-Hurtado et al., 2012):

$$\boldsymbol{g}_{k,l} = \beta_{l,k} \boldsymbol{e}_p(l,k)^{(\alpha_{l,k}-1)}. \tag{41}$$

The factors α and β are computed as

$$\beta_{l,k} = \max(\boldsymbol{e}_{pmax}(k))^{(1-\alpha_{k,l})} \tag{42}$$

$$\alpha_{k,l} = 1 - \log K / \log M_{l,k}\,, \tag{43}$$

where $M_{l,k}$ is the SNR at k-th sub-band and l-th frame, and $\boldsymbol{e}_{pmax}(k)$, a vector that holds the maximum values of the primary envelopes, is obtained from the previous N_{max} frames:

$$\boldsymbol{e}_{pmax}(k) = [\max(\boldsymbol{e}_p(l,k)) \ldots \max(\boldsymbol{e}_p(l - N_{max}, k))] \tag{44}$$

To avoid computational overflow and preserve the binaural cues, the value of α is constrained in the range $\alpha = [0, 5]$. To minimize artifacts and achieve better quality outputs, the history stored in the vector $\boldsymbol{e}_{pmax}$ should hold at least one second, but two-seconds memory, i.e. $N_{max} = \lceil 2 f_s / N \rceil$, is recommended. Since α and β are fixed for a given frame, these gains can also be computed in the sample-by-sample basis.

To estimate the SNR at the given sub-band and frame, the signal and noise power are obtained from the envelopes of the primary and secondary channel. This approach reduces miss-classification errors in the SNR estimation when the input SNR is low. To obtain a reliable noise estimate, the noise power is updated using a rule derived from the noise PSD estimator proposed in (Ris & Dupont, 2001):

$$\begin{aligned}
&P_e = \|e_s(l,k)\|^2 \\
&if\ |P_e - P_v(l-1,k)| < \epsilon\sqrt{\sigma_v(l-1,k)} \\
&\quad P_v(l,k) = \lambda_v P_v(l-1,k) + (1-\lambda_v) P_e \\
&\quad \sigma_v(l,k) = \delta\sigma_v(l-1,k) + (1-\delta)\,|P_e - P_v(l-1,k)|^2 \\
&else \\
&\quad P_v(l,k) = P_v(l-1,k) \\
&\quad \sigma_v(l,k) = \sigma_v(l-1,k) \\
&end
\end{aligned} \tag{45}$$

where $P_v(l,k)$ is the noise power at the k-th sub-band and l-th frame, $\sigma_v(l,k)$ is an estimate of the variance of P_v, λ and δ are time constants to smooth the estimation, and ϵ is a threshold coefficient. Finally, the frame SNR is estimated by

$$M_{l,k} = \max\left(\frac{P_x(l,k)}{P_v(l,k)} - 1\,,\, 1\right) \tag{46}$$

where P_x is the power of the primary channel estimated by

$$P_x(l,k) = \lambda_x P_x(l-1,k) + (1-\lambda_x)\left\|e_p(l,k)\right\|^2 \tag{47}$$

The values $\lambda_v = 0.95$, $\lambda_x = 0.9$, $\delta = 0.9$, and $\epsilon = 5$ are selected in (Marin-Hurtado et al., 2012) to achieve good performance.

The performance of the BSS-PP depends on the tuning of two parameters: K and N. Whereas K controls the expansion of the dynamic range, N defines how often the parameters to compute the noise-suppression gains are updated. A detailed analysis of the effect of these parameters on the SNR improvement and sound quality is presented in (Marin-Hurtado et al., 2012). In summary, $K = 0.01$ and $N = 8192$ show to be suitable for all scenarios. The mathematical proof that localization cues are preserved in the BSS-PP algorithm is included in (Marin-Hurtado et al., 2012).

3.5.1 Advantages and limitations

In the BSS-PP method, the noise-suppression gains are computed to expand the dynamic range of the noisy signal, in such a way that the maximum signal level is maintained while the noise level is pushed down. The maximum signal level is estimated from the primary channel, and the noise level from the secondary channel. Theoretical analysis conducted in (Takahashi et al., 2009) show that ICA-based BSS algorithms such as the algorithm used in the BSS-PP method provides an accurate noise estimate under non-point-source noise scenarios (e.g., diffusive or babble noise). Therefore, the performance of this method under these scenarios is expected to be high. Since BSS-PP tracks the envelopes of the target speech and noise level simultaneously, it is expected a good performance under highly non-stationary environments. On the other hand, when the interfering signals are few point sources, the BSS algorithm can provide accurate noise estimation only if the target signal is dominant. Thus, the performance of the BSS-PP algorithm is expected to be low under these scenarios at very low input SNR. Fortunately, these kind of scenarios are uncommon. All the above statements are verified through experiments discussed in the next section. In general, the BSS-PP method shows to be efficient in the removal of background noise, provides an acceptable speech quality, preserves the localization cues for both target and interfering signals, and outperforms existing BSS-based methods in terms of SNR improvement and noise reduction (Marin-Hurtado et al., 2012).

4. Comparative study

This chapter discussed different methods to preserve the localization cues in a binaural noise-reduction system based on BSS. These methods are summarized in the Table 1. Based on common features of the algorithms, these methods can be classified in three categories: BSS

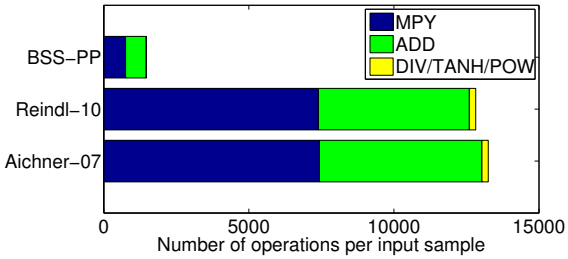

Fig. 9. Number of operations for BSS-PP, Reindl-10, and Aichner-07 per input sample grouped into additions (ADD), multiplications (MPY), divisions (DIV), hyperbolic tangent (TANH), and power raise (POW).

constrained optimization, spatial-placement filters, and BSS post processing to enhance the unprocessed signals. In the first category, the BSS filter weights, w_{qp}, are designed to perform source separation as well as to preserve the localization cues. In the second category, the BSS output corresponding to the estimate of the target signal is enhanced by a FIR filter that restores the localization cues. In the third category, the BSS outputs are used to compute noise-suppression gains that enhance the unprocessed signals. Under the third category, we can include the post-processing methods based on adaptive filters, Wiener filter, and perceptually-inspired processing.

Different reports have been shown that methods based on BSS constrained optimization and spatial-placement filters are unable to provide simultaneous preservation of localization cues for the target and interfering signals. In addition, most of these methods perform a mapping of the direction of arrival of the interfering signals to the direction of arrival of the target signal, which may be perceptually annoying. On the contrary, most methods belonging to the third category, BSS post processing to enhance the unprocessed signals, can preserve the localization cues for both target and interfering signals simultaneously under certain conditions. In particular, among the different methods analyzed, the BSS post-processing method based on Wiener filter and the perceptually-inspired post processing are the only methods able to preserve these localization cues simultaneously.

Since the gains and outputs are computed in the sample-by-sample basis, the processing delay is very small ($<$ 1 ms) in the BSS-PP method compared to other BSS-based post-processing methods such as the method based on adaptive filters, Aichner-07, (Section 3.3), and the method based on Wiener filter, Reindl-10, (Section 3.4). In Aichner-07 and Reindl-10 methods, the processing delay is around 6 ms. In addition, the computational complexity of BSS-PP is significantly smaller than Aichner-07 and Reindl-10 (Fig. 9).

4.1 Experiment

Among the different methods discussed in this chapter, only Aichner-07, Reindl-10, and BSS-PP are evaluated in this experiment. This selection takes into account only the BSS post-processing methods capable of preserving the localization cues for the target and interfering signals simultaneously under certain environmental conditions (Table 1). These methods are implemented in Matlab and tested under different scenarios. Simulations to discern the performance of these techniques are conducted under the following scenarios:

Method	Strategy	Preserve Target Cues	Preserve Noise Cues	Comp. Cost	Processing Delay	Ref.
Takatani-05	BSS constrained optimization	Yes	No	High	?	(Takatani et al., 2005)
Aichner-07B	BSS constrained optimization	Yes	No	High	?	(Aichner et al., 2007)
Wehr-06A	Spatial-placement filter	No	Mapped to target DoA	Medium	?	(Wehr et al., 2006)
Wehr-06B	Spatial-placement filter	Yes	Mapped to target DoA	Medium	?	(Wehr et al., 2006)
Aichner-07	Post processing based on adaptive filters	Yes	Under certain conditions	Medium	$\sim$ 6 ms	(Aichner et al., 2007)
Reindl-10	Post processing based on Wiener filter	Yes	Yes	Medium	$\sim$ 6 ms	(Reindl et al., 2010)
BSS-PP	Perceptually-inspired post processing	Yes	Yes	Low	$\sim$ 1 ms	(Marin-Hurtado et al., 2012)

Table 1. Summary of the binaural noise-reduction methods based on BSS. Processing delay is estimated for a system working at 16 kHz sampling frequency. Question mark is included for the methods not analyzed in the comparative study.

1. Single source under constant-SNR diffusive noise. This scenario is widely used to test virtually all binaural noise-reduction techniques. This background noise is generated by playing uncorrelated pink noise sources simultaneously at 18 different spatial locations.
2. Single source under babble (or cafeteria) noise. The background noise corresponds to a real recording in a cafeteria.
3. Multi-talker. In this scenario, four distinguishable speakers are placed at different azimuthal positions: 40^o, 80^o, 200^o and 260^o.

The above scenarios are generated by filtering the target signal with the HRTF measured for a KEMAR manikin in absence of reverberation (Gardner & Martin, 1994). The target signal is placed at eight different azimuthal angles: 0^o, 30^o, 90^o, 120^o, 180^o, 240^o, 270^o and 330^o, where 0^o corresponds to the front of the KEMAR, 90^o corresponds to the right ear, and 270^o to the left ear. Target signals are speech recordings of ten different speakers and sentences taken from the IEEE sentence database (IEEE Subcommittee, 1969). For all scenarios, the interfering signals are added to the target signal at different SNR.

Since the HRTF database in (Gardner & Martin, 1994) is for non-reverberant environments, a secondary database using reverberant conditions is created using the HRTF recordings described in (Jeub et al., 2009; RWTH Aachen University, 2010). This database is included since it is widely known that the performance of the majority of the noise-reduction algorithms is degraded significantly when reverberation is present. This database assumes a babble noise scenario and the following rooms: studio ($RT_{60} = 0.12$s), meeting room ($RT_{60} = 0.23$s), office ($RT_{60} = 0.43$s), and lecture room ($RT_{60} = 0.78$s).

The performance of these techniques is analyzed using the broadband intelligibility-weighted SNR improvement (ΔSNR-SII) (Greenberg et al., 1993). For the subjective test, a MUSHRA (multiple stimulus test with hidden reference and anchor) test is used to assess the overall sound quality. The protocol in (ITU-R, 2003) is used for the subjective test.

4.2 Performance evaluation

SNR improvement for diffusive, babble, and multi-talker scenarios is plotted in Figures 10-12. In general, the perceptually-inspired post-processing method (BSS-PP) outperforms the other BSS-based noise-reduction methods in most scenarios.

The poor performance of BSS-PP in the multi-talker scenario at low input SNR is explained by the errors introduced by a wrong selection of the primary output. When an ideal output selection algorithm is used (dashed line in Fig. 12), the performance of BSS-PP is similar or better than that of the other BSS-based methods. The output selection algorithm can be made more robust by using a direction-of-arrival-estimation algorithm or a permutation algorithm at expenses of increasing the computational complexity. However, scenarios with very few interfering signals at input SNR < 0 dB such as the multi-talker scenario of Fig. 12 are very uncommon, and they are not challenging for the auditory system without any hearing aid. Likewise, binaural noise-reduction methods are useful for challenging scenarios such as babble noise at low input SNR. Since BSS-PP provides an excellent performance under these scenarios (Fig. 11), the output-selection algorithm used by BSS-PP is enough for a large set of practical applications.

Up to this point the performance of all methods has been verified under non-reverberant scenarios. For reverberant scenarios, Fig. 13 shows that for a large reverberant room

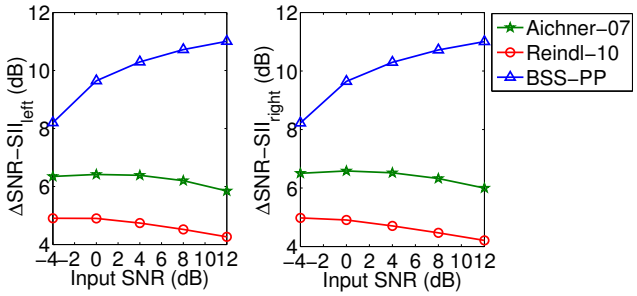

Fig. 10. SNR improvement under diffusive noise scenario.

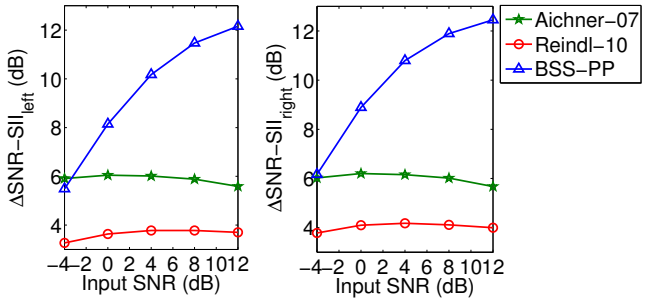

Fig. 11. SNR improvement under babble noise scenario.

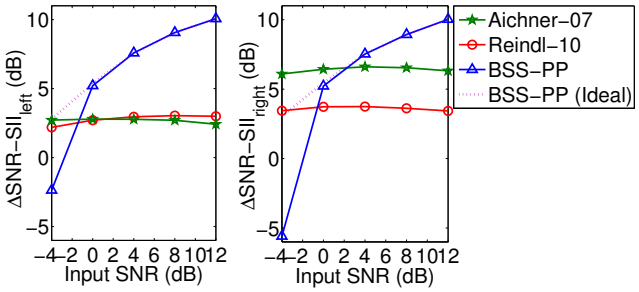

Fig. 12. SNR improvement under multi-talker scenario. The dashed line is the performance for an ideal output-selection algorithm.

($RT_{60} = 0.78$s), BSS-PP provides an acceptable SNR improvement and outperforms the other existing methods for input SNR ≥ 0 dB. Results for other reverberant rooms are included in (Marin-Hurtado et al., 2012).

A subjective test is conducted to assess the subjective sound quality of the methods under study. These results are summarized in the Fig. 14. Sound quality is graded in the scale $[0, 100]$, with 100 the highest value corresponding to a clean signal. To perform the grading, the subject listened to the samples that included clean speech, unprocessed speech in babble noise at an input SNR of 0 dB, and enhanced speech processed by Aichner-07, Reindl-10, and BSS-PP methods. The reference and hidden reference signals are unprocessed noisy

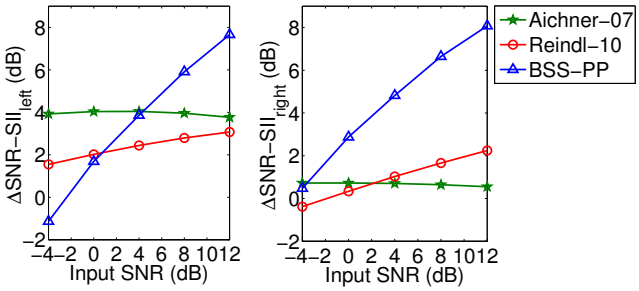

Fig. 13. SNR improvement under babble noise scenario in a lecture room (reverberant condition $RT_{60} = 0.78$s).

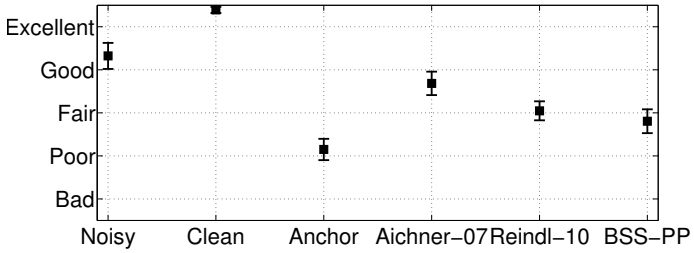

Fig. 14. Subjective test results for speech quality. Reference: speech in babble noise; anchor: noisy speech distorted according to (ITU-R, 2003).

speech while the anchor signal is noisy speech distorted according to (ITU-R, 2003). All samples are five-seconds long, and they are presented randomly to the subject. A total of 20 normal-hearing subjects participated in the experiment. Results show that there is a distortion in the speech quality for all methods, and a subject preference for the speech quality of the unprocessed noisy signal. The methods providing the lowest noise reduction (Aichner-07 and Reindl-10) achieved the best speech quality, and the methods with the highest noise reduction (BSS-PP), the lowest speech quality. However, the speech quality of BSS-PP is higher than the speech quality of the anchor signal (an artificially-distorted speech signal).

5. Conclusions

This chapter described different binaural BSS-based noise-reduction algorithms that are promising for the reduction of the background noise and the preservation of the direction of arrival of the target and interfering signals. The preservation of the direction of arrival, also known as localization cues, is an important issue for some applications such as binaural hearing aids. In these devices, the displacement or lost of these localization cues is reported as perceptually annoying by hearing-impaired users.

The methods reported in the literature to preserve the localization cues in a binaural BSS-based noise-reduction algorithm can be classified into three categories: a) BSS algorithms based on contrained optimization to preserve the localization cues (Section 3.1); b) restoration of the localization cues by means of post processing applied to the BSS output related to the target signal (e.g., spatial-placement filter on Section 3.2); and c) enhancement of the

unprocessed inputs by noise-reduction gains computed from the BSS outputs (e.g., adaptive filters, Wiener filter, and perceptual post-processing methods described on Sections 3.3, 3.4, and 3.5). All methods proposed in the literature can preserve the localization cues for the target signal. However, the methods belonging to the first and second category, BSS constrained optimization and spatial-placement filters, cannot preserve the localization cues for the interfering signals. In most cases, these localization cues are mapped to the direction of arrival of the target signal, which suggests that these algorithms are not practical for binaural hearing aids. On the contrary, binaural BSS-based noise-reduction algorithms belonging to the third category, i.e., those methods that compute noise-suppression gains from the BSS outputs and apply these gains to the unprocessed signals, can preserve the localization cues for both target and interfering signals simultaneously. This preservation is identified through subjective and theoretical analysis. This chapter described three methods belonging to the third category: post processing with adaptive filters (Aichner-07), post processing with Wiener filter (Reindl-10), and perceptually-inspired post processing (BSS-PP). An experimental evidence, confirmed through mathematical analysis, showed the post processing based on adaptive filters (Aichner-07) works only in the determined case, i.e., when the number of source signals is equal o lower than the number of sensors. On the contrary, the methods based on Wiener-filter post processing (Reindl-10) and perceptually-inspired post processing (BSS-PP) preserve the localization cues even in the undetermined case.

A comparative study conducted with the Aichner-07, Reindl-10, and BSS-PP methods under different environments showed that BSS-PP outperforms the other methods in terms of SNR improvement and noise reduction. In addition, the BSS-PP method provides a significant reduction in the number of operations compared to the other two methods, and its processing delay is very small. Hence, the BSS-PP turns out a feasible solution for a binaural hearing aid. However, there are two limitations in the BSS-PP method. First, the subjective sound quality is acceptable, with a subjective sound quality graded slightly below the subjective sound quality of the Aichner-07 and Reindl-10 methods. Second, the BSS algorithm demands wireless transmission at full rate. This issue is also present in the Aichner-07 and Reindl-10 methods.

6. Future work

Although the BSS-PP method is a promising binaural noise-reduction algorithm, it is necessary to solve two issues to obtain a practical implementation for a binaural hearing aid. First, to improve the sound quality, the dynamic range expansion performed by the post processing stage must include additional information to take into account a sound quality criteria, or use another perceptual model. Second, to reduce the transmission bandwidth, it is necessary to develop distributive or reduced bandwidth BSS algorithms, or to employ strategies other than BSS to estimate the target and interfering signals.

Most processing in the BSS-PP method can be easily replaced by an analog processing except the BSS algorithm. A mixed-signal solution may reduce computational complexity and power consumption. To obtain a full-analog solution, analog BSS algorithms have to be developed.

Although most BSS-based noise-reduction algorithms such as Reindl-10 and BSS-PP were not initially designed to deal with reverberant conditions, their performance under these environments is acceptable. Hence, their performance could be improvement by modifications in the mathematical framework to take into account the effect of reverberation.

Finally, it is known that the speech intelligibility in noise-reduction applications can be improved by applying a binary mask to the unprocessed signal (Loizou & Kim, 2011). Hence, binary masking can be combined with a BSS algorithm in order to obtain a source separation algorithm that reduces the background noise and improves the speech intelligibility simultaneously. Although some attempts have been explored in (Han et al., 2009; Jan et al., 2011; Mori et al., 2007; Takafuji et al., 2008), these methods are unable to preserve the localization cues for both target and interfering signals simultaneously. Hence, it is necessary to develop post processing algorithms to preserve the localization cues in BSS-based binary masking algorithms.

7. Acknowledges

This work is supported in part by Texas Instruments Inc., formerly National Semiconductor Corporation. Jorge Marin wants to thank Georgia Institute of Technology–USA, Universidad del Quindio–Colombia and Colciencias–Colombia for their financial support.

8. References

Aichner, R., Buchner, H., & Kellermann, W. (2006). A novel normalization and regularization scheme for broadband convolutive blind source separation, *Proc. Int. Symp. Independent Component Analysis and Blind Signal Separation, ICA 2006*, Vol. 3889, pp. 527–535.

Aichner, R., Buchner, H., Zourub, M. & Kellermann, W. (2007). Multi-channel source separation preserving spatial information, *Proc. IEEE Int. Conf. Acoust., Speech Signal Process., ICASSP 2007*, Vol. 1, pp. I–5–I–8.

Gardner, B. & Martin, K. (1994). HRTF measurements of a KEMAR dummy-head microphone, *Technical Report 280*, MIT Media Lab Perceptual Computing. http://sound.media.mit.edu/KEMAR.html.

Greenberg, J. E., Peterson, P. M. & Zurek, P. M. (1993). Intelligibility-weighted measures of speech-to-interference ratio and speech system performance, *J. Acoust. Soc. Amer.* 94(5): 3009–3010.

Han, S., Cui, J. & Li, P. (2009). Post-processing for frequency-domain blind source separation in hearing aids, *Proc. Int. Conf. on Information, Communications and Signal Processing, 2009. ICICS 2009*, pp. 1 –5.

Haykin, S. (2000). *Unsupervised Adaptive Filtering*, Vol. 1: Blind Source Separation, John Wiley and Sons.

IEEE Subcommittee (1969). IEEE recommended practice for speech quality measurements, *IEEE Trans. Audio Electroacoust.* pp. 225–246.

ITU-R (2003). Recommendation BS.1534-1: Method for the subjective assessment of intermediate quality levels of coding systems.

Jan, T., Wang, W. & Wang, D. (2011). A multistage approach to blind separation of convolutive speech mixtures, *Speech Communication* 53(4): 524 – 539.

Jeub, M., Schafer, M. & Vary, P. (2009). A binaural room impulse response database for the evaluation of dereverberation algorithms, *Proc. Int. Conf. Digital Signal Process.*, pp. 1 –5.

Kocinski, J. (2008). Speech intelligibility improvement using convolutive blind source separation assisted by denoising algorithms, *Speech Commun.* 50(1): 29 – 37.

Loizou, P. & Kim, G. (2011). Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions, *IEEE Transactions on Audio, Speech, and Language Processing* 19(1): 47 –56.

Marin-Hurtado, J. I., Parikh, D. N. & Anderson, D. V. (2011). Binaural noise-reduction method based on blind source separation and perceptual post processing, *Proc. Interspeech 2011*, Vol. 1, Florence, Italy, pp. 217–220.

Marin-Hurtado, J. I., Parkih, D. N. & Anderson, D. V. (2012). Perceptually inspired noise-reduction method for binaural hearing aids, *IEEE Transactions on Audio, Speech and Language Processing* 20(4): 1372–1382.

Moore, B. C. J. (2007). Binaural sharing of audio signals: Prospective benefits and limitations, *The Hearing Journal* 60(11): 46–48.

Mori, Y., Takatani, T., Saruwatari, H., Shikano, K., Hiekata, T. & Morita, T. (2007). High-presence hearing-aid system using dsp-based real-time blind source separation module, *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process., 2007. ICASSP 2007*, Vol. 4, pp. IV–609 –IV–612.

Noohi, T. & Kahaei, M. (2010). Residual cross-talk suppression for convolutive blind source separation, *Proc. Int. Conf. Comp. Eng. Technology (ICCET)*, Vol. 1, pp. V1–543 –V1–547.

Parikh, D., Ikram, M. & Anderson, D. (2010). Implementation of blind source separation and a post-processing algorithm for noise suppression in cell-phone applications, *Proc. IEEE Int. Conf. Acoust., Speech Signal Process., ICASSP*, pp. 1634 –1637.

Parikh, D. N. & Anderson, D. V. (2011). Blind source separation with perceptual post processing, *Proc. IEEE 2011 DSP/SPE Workshop*.

Park, K. S., Park, J., Son, K. & Kim, H. T. (2006). Postprocessing with wiener filtering technique for reducing residual crosstalk in blind source separation, *Signal Processing Letters, IEEE* 13(12): 749 –751.

Reindl, K., Zheng, Y. & Kellermann, W. (2010). Speech enhancement for binaural hearing aids based on blind source separation, *Proc. Int. Symp. Commun. Control Signal Process. (ISCCSP)*, pp. 1 –6.

Ris, C. & Dupont, S. (2001). Assessing local noise level estimation methods: Application to noise robust ASR, *Speech Commun.* 34(1-2): 141 – 158.

RWTH Aachen University (2010). Aachen impulse response (AIR) database - version 1.2. http://www.ind.rwth-aachen.de/AIR.

Smith, P., Davis, A., Day, J., Unwin, S., Day, G. & Chalupper, J. (2008). Real-world preferences for linked bilateral processing, *The Hearing Journal* 61(7): 33–38.

Sockalingam, R., Holmberg, M., Eneroth, K. & Shulte, M. (2009). Binaural hearing aid communication shown to improve sound quality and localization, *The Hearing Journal* 62(10): 46–47.

Takafuji, R., Mori, Y., Saruwatari, H. & Shikano, K. (2008). Binaural hearing-aid system using simo-model-based ica and directivity-dependency-reduced binary masking, *Proc.* 9th *Int. Conf. Signal Process., ICSP 2008*, pp. 320 –323.

Takahashi, Y., Takatani, T., Osako, K., Saruwatari, H. & Shikano, K. (2009). Blind spatial subtraction array for speech enhancement in noisy environment, *IEEE Transactions on Audio, Speech and Language Processing* 17(4): 650 –664.

Takatani, T., Ukai, S., Nishikawa, T., Saruwatari, H. & Shikano, K. (2005). Evaluation of SIMO separation methods for blind decomposition of binaural mixed signals, *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*, pp. 233–236.

Van den Bogaert, T., Klasen, T. J., Moonen, M., Van Deun, L. & Wouters, J. (2006). Horizontal localization with bilateral hearing aids: Without is better than with, *J. Acoust. Soc. Amer.* 119(1): 515–526.

Wehr, S., Puder, H. & Kellermann, W. (2008). Blind source separation and binaural reproduction with hearing aids: An overview, *Proc. ITG Conf. Voice Communication (SprachKommunikation)* pp. 1 –4.

Wehr, S., Zourub, M., Aichner, R. & Kellermann, W. (2006). Post-processing for BSS algorithms to recover spatial cues, *Proc. Int. Workshop Acoust. Echo Noise Control (IWAENC).*