

# A Noise-Reduction Method With Coherence Enhancement for Binaural Hearing Aids

Johnny Werner and Márcio Holsbach Costa

**Abstract**—In this work, a variation of the multichannel Wiener filter (MWF) for noise reduction in binaural hearing aid applications is proposed. This method provides improved spatial preservation for acoustic scenarios comprised of one single-point target (speech) and one single-point interfering (noise) acoustic sources. It employs a regularization penalty term based on the inverse of the magnitude of the interaural coherence (IC), with the aim of enhancing the original azimuth perception of the interferent source. The proposed penalty term artificially intensifies the IC of the residual interference, increasing similarity in both ears. This is of special value for low-coherent reverberant acoustic signals, in which the reliability of the received binaural cues associated to the single-point interferent source was degraded by multiple acoustic reflections. Simulation results obtained with objective criteria show that, in addition of providing improved spatial preservation for the interferent source, the proposed approach may also provide higher noise reduction performance as compared to the conventional MWF approach. Further, it also yields an extended range for the target and interference spatial-preservation trade-off, as compared to a previously developed competing method. Psychoacoustic experiments with normal hearing volunteers corroborate theoretical and simulation findings.

**Index Terms**—Binaural hearing aids, noise reduction, binaural cues, interaural coherence, Wiener filter.

## I. INTRODUCTION

**H**UMAN hearing losses may result in social isolation, professional difficulties and risk to personal safety. Hearing aids are devices designed for compensating hearing impairment. However, it is well known that their performance and acceptance may be highly affected by secondary factors, such as interfering noise [1], which does not only decrease intelligibility and acoustic comfort, but may also reduce the daily use of the device by the user [2].

In this sense, modern digital hearing aids are composed of not only a hearing compensation strategy, but also many other subsystems, such as: noise reduction, feedback cancellation [3], and dereverberation [4]. However, it is well known that dissimilar modifications of the original signals presented to the ears may affect the binaural cues, having the potential to distort the ability of the human auditory system in perceiving

the direction of arrival of acoustic sources [5] [6].

Keeping the spatial perception of the original acoustic scenario is of paramount importance for safety and environment awareness of hearing aid users. When some binaural cues are significantly distorted, hearing aid users have difficulties in accurately identifying the origin of noise sources of interest, such as alarms, horns or beeps. Therefore, signal-processing methods for hearing aids that do not distort the perception of the acoustic scenario are of great interest.

According to [1], more than 80% of the hearing-impaired people have both ears affected by a reduction in hearing ability, requiring the simultaneous use of two devices. For these cases, bilateral hearing aids<sup>1</sup> may not adequately preserve the original acoustic scenario, distorting the ability to locate, separate, and track sound sources [6]. On the other hand, binaural hearing aids<sup>2</sup> have the advantage of exploiting the spatial information that reaches both ears [1], showing improved performance for preservation of the acoustic cues [7].

Psychoacoustic experiments have shown that localization in the azimuth plane is essentially based on the interaural time difference (ITD), which measures the transmission time difference between both ears; and the interaural level difference (ILD), which measures the magnitude difference between signals in the ears [6]. According to the duplex theory, the ITD is the most relevant binaural cue at frequencies below 1.5 kHz, while ILD is more relevant above that [8]. These two binaural cues are created mostly by the human head, which diffracts acoustic waves that reach both ears (head shadow effect) [9]. Thus, received sounds are asymmetrically attenuated and delayed, generating differences between sounds in both ears. This feature provides the listener with the necessary information to predict the location of the acoustic sources and reconstruct the acoustic scene. This also helps the listener to naturally focus on a particular sound source and disregard the unwanted ones. This phenomenon is called spatial release from masking effect [10].

In reverberant scenarios, there are acoustic reflections coming from various directions. These overlapped received sounds may generate extraneous ITDs and ILDs that may not correspond to the true position of the acoustic sources, degrading the performance of the localization process [8] [11] [12] [14] [15]. Despite normal healthy listeners are naturally able to deal with the effect of reflections [16], the source

This work was supported by CAPES and CNPq under Grant 315020/2018-0.

J. Werner is with the Departamento Acadêmico de Elétrica, Universidade Tecnológica Federal do Paraná (UTFPR), Pato Branco, zip 85503-390, Brazil (e-mail: werner@utfpr.edu.br). M. H. Costa is with the Departamento de Engenharia Elétrica e Eletrônica, Universidade Federal de Santa Catarina (UFSC), Florianópolis, zip 88040-900, Brazil (e-mail: costa@eel.ufsc.br).

Digital Object Identifier: 10.14209/jcis.2020.34

<sup>1</sup> Received signals are processed independently in the left and right hearing aids.

<sup>2</sup> Control parameters and signals received in the right and left hearing aids are mutually shared through a wireless link.

localization ability of hearing aid users may be strongly affected [11]. Hence, hearing aid devices should be carefully designed to work in such reverberant acoustic scenarios.

The binaural MWF [19] is a popular noise reduction method widely applied in the binaural hearing aid context. The MWF has acceptable computational complexity, high robustness and allows easy integration with other speech processing methods [20]. As major advantages, it provides significant increase in the signal-to-noise ratio (SNR) and naturally preserves the original binaural cues of the target source (speech) [21]. In contrast, the binaural cues of the residual noise are frequently distorted in a way that noise sources are perceived as having the same azimuth as the target source, changing the perception of the original acoustic scenario. Thus, after processing the input signals by the binaural MWF, there is no longer spatial separation between the processed speech and residual noise, and no binaural unmasking can be exploited by the human auditory system [22].

In attempt to overcome this drawback, different MWF-based strategies for noise reduction with spatial preservation have been proposed [23] [24] [25] [26]. Mainly, they apply auxiliary terms to the MWF cost function to penalize solutions that do not preserve the original binaural cues of the residual noise [27]. Hence, a trade-off between noise reduction and spatial preservation is commonly provided. In this context, some attempts to preserve ITD and ILD of the residual noise were proposed in [24] and in [28]. In [23], the MWF-ITD was introduced for preserving both speech and noise ITD. However, extensive experiments carried in [27] indicated that the MWF-ITD does not provide perceptually relevant results, showing that preserving only the ITD is not sufficient to avoid distortion of the original acoustic scene.

The interaural coherence (IC) has been found to be an important measure of similarity between acoustic signals in both ears. In complex listening situations, in which the signals received by the ears of the listener are degraded by acoustic reflections, the coherence tends to decrease. In [29], it was proposed a method for binaural dereverberation that employs the IC. In this method, online statistics of the coherence are used to set individual filter gains to attenuate time-frequency tiles in which the IC is small (due to reverberation) and preserve regions with large IC (in which the direct sound is predominant).

The IC is also important in defining the reliability of acoustic binaural cues such as ITD and ILD [11] [17]. For small coherence values, ITD tends to lose its meaning, since there are no waveforms with common characteristics that can be compared [12]. Thus, with the purpose of locating acoustic sources, the authors in [11] proposed to select only the ITD and ILD binaural cues associated to large coherence values, which were similar to those obtained in free-field.

In [12], the importance of IC for effective ITD preservation was analyzed. The authors assessed the lateralization perception of volunteers in reverberant environments for different values of the IC magnitude of the input noise ( $|C_{in}^v|$ ). As a result, approximately 50% of hemisphere inversions were observed for mid-band noise (range from 630 Hz to 800 Hz) when  $|C_{in}^v| = 0.2$ , whereas no inversions were reported for  $|C_{in}^v| > 0.8$ . Thus, the authors in [12] concluded that the IC should be considered cooperatively with the ITD to improve

the localization of acoustic sources. Later, in [18], it was shown that IC is a nonlinear function of ITD.

The association between the Wiener filter and the IC was previously employed in [30] for dereverberation, and was further refined in [31] and in [6], in which the head shadow effect was also considered. It was also used for noise reduction with spatial preservation for diffuse noise in [26]. This technique, named MWF-IC, provides noise reduction filters, which minimize IC differences between input (received at the microphones) and output (processed) signals. Recently, in a work from our group [27], it was demonstrated that the MWF-IC [26] can be also applied for preserving the ITD of a single-point acoustic noise source in free-field.

In this paper, following the findings presented in [11] and [12] about the intrinsic relation between IC and the correct localization of acoustic sources, we propose a new MWF based method, for both noise reduction and spatial preservation of one speech and one interferent single-point sources immersed in a reverberant acoustic scenario. This new method is called Coherence-based MWF (MWF-CB), and consists of a variation of the conventional binaural MWF method, which artificially intensifies the magnitude of the coherence of the processed noise at the output loudspeakers of the hearing aids.

A preliminary version of this method was first presented in [13]. Here, it is further explored. Psychoacoustic experiments are formally conducted, in which the localization performance of the proposed method is evaluated. Following, it is now demonstrated that the proposed method is efficient in preserving the true azimuth of the noise source, as compared to a similar concurrent method, without losing the original SNR improvement capability of the binaural MWF. In addition, the new results also comprise a wider variety of reverberant acoustic scenes with more realistic input SNR range, in which various set points are explored. Finally, a deeper theoretical basis and a new mathematical point of view, as well as improved discussion, are provided.

The main contributions of this work are: (a) an efficient method for noise reduction and spatial preservation of a reverberant acoustic scenario in binaural hearing aids is proposed; (b) its interpretation as a constrained optimization problem is presented; (c) an extensive assessment and comparison performance, using objective criteria, against the conventional binaural MWF [19] and MWF-IC [26] methods is presented; and (d) psychoacoustic experiments with volunteers are provided to show the lateralization accuracy of the proposed technique.

In this document, matrices are represented by bold uppercase letters; vectors are represented by bold lowercase letters, and scalars are represented as italics.

## II. BINAURAL HEARING AID SYSTEM

The application context of this work comprises a binaural fitting of hearing aids working in full-duplex mode without bit-rate limitations. Fig. 1 shows the binaural setup, in which each hearing aid is equipped with  $M$  microphones. The operating scenario assumes the existence of one speech source of interest,  $x(t)$ , and one single-point interfering noise source,  $v(t)$ , in a reverberant room. It is assumed that all sources have

a fixed (or slowly varying) position in a given time-window. For the incoming signals, it is applied the frequency domain decomposition through an  $N$ -bin Short-Time Fourier Transform (STFT). For a sampling frequency of  $f_s$  samples per second, for each time-frame  $\lambda$  and each frequency  $k$ , the received signals are defined as

$$y_{L,l}(\lambda, k) = x_{L,l}(\lambda, k) + v_{L,l}(\lambda, k), \quad (1)$$

$$y_{R,r}(\lambda, k) = x_{R,r}(\lambda, k) + v_{R,r}(\lambda, k), \quad (2)$$

in which  $y_{L,m}(\lambda, k)$  with  $m = \{1, \dots, M\}$  are the STFTs of the signals acquired at the  $M$  microphones at the left (L) hearing aid, and  $y_{R,m}(\lambda, k)$  are the signals at the right (R) hearing aid. The received signals are stacked in the  $2M \times 1$  vector as

$$\mathbf{y}(\lambda, k) = \mathbf{x}(\lambda, k) + \mathbf{v}(\lambda, k), \quad (3)$$

in which  $\mathbf{y}(\lambda, k) = [y_{L,1}(\lambda, k) \dots y_{L,M}(\lambda, k) y_{R,1}(\lambda, k) \dots y_{R,M}(\lambda, k)]^T$  is the noisy-speech vector;  $\mathbf{x}(\lambda, k) = [x_{L,1}(\lambda, k) \dots x_{L,M}(\lambda, k) x_{R,1}(\lambda, k) \dots x_{R,M}(\lambda, k)]^T$  is the speech component;  $\mathbf{v}(\lambda, k) = [v_{L,1}(\lambda, k) \dots v_{L,M}(\lambda, k) v_{R,1}(\lambda, k) \dots v_{R,M}(\lambda, k)]^T$  is the interferent noise vector; and  $(\cdot)^T$  represents the transpose operator.

For a clean notation, the time-frame  $\lambda$  and  $k$  frequency indices will be omitted in the following, wherever possible.

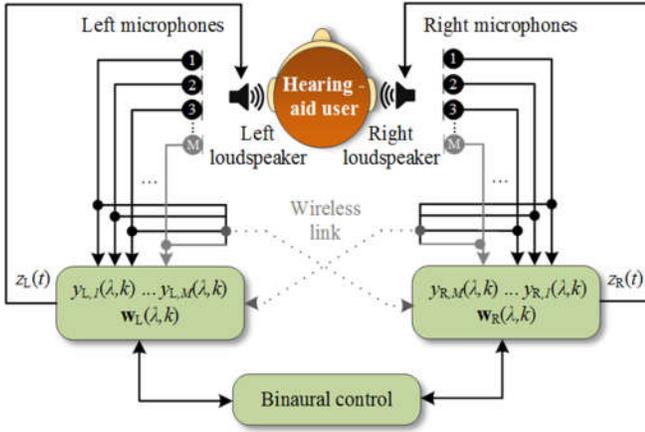


Fig. 1. Binaural hearing aid setup.

### III. MULTICHANNEL WIENER FILTER

The cost function of the binaural MWF [19], for each bin, is defined as

$$J_{WF}(\mathbf{w}) = E\{|x_L - \mathbf{w}_L^H \mathbf{y}|^2\} + E\{|x_R - \mathbf{w}_R^H \mathbf{y}|^2\}, \quad (4)$$

in which  $E\{\cdot\}$  indicates the expected value operator;  $|\cdot|$  is the absolute value;  $(\cdot)^H$  is the conjugate transpose operator;  $x_L$  and  $x_R$  are the frequency domain representation of the desired (speech) signal component at the reference microphone, respectively, in the left and right hearing aids; and  $\mathbf{w}_L$  and  $\mathbf{w}_R$  are, respectively, the left and right coefficient vectors of the MWF, both with dimension  $2M \times 1$ . The complex stacked weight vector  $\mathbf{w}$ , with dimension  $4M \times 1$ , is defined as  $\mathbf{w} = [\mathbf{w}_L^T \mathbf{w}_R^T]^T$ . Manipulating (4) leads to [27]

$$J_{WF}(\mathbf{w}) = \mathbf{q}^T \Theta_x \mathbf{q} - \mathbf{q}^T \Theta_x \mathbf{w} - \mathbf{w}^H \Theta_x \mathbf{q} + \mathbf{w}^H \Theta_y \mathbf{w}, \quad (5)$$

in which  $\mathbf{q} = [\mathbf{q}_L^T \mathbf{q}_R^T]^T$ ; the deterministic vectors  $\mathbf{q}_L$  and  $\mathbf{q}_R$ , both with dimensions  $2M \times 1$ , contain 1 in the element corresponding to the respective (left/right) reference microphone and zeros otherwise;

$$\Theta_x = \begin{bmatrix} \Phi_x & \mathbf{0} \\ \mathbf{0} & \Phi_x \end{bmatrix}, \quad \Theta_y = \begin{bmatrix} \Phi_y & \mathbf{0} \\ \mathbf{0} & \Phi_y \end{bmatrix}; \quad (6)$$

$\mathbf{0}$  is a  $2M \times 2M$  null matrix, containing zeros;  $\Phi_x = E\{\mathbf{x}\mathbf{x}^H\}$  is the  $2M \times 2M$  coherence matrix of the speech; and  $\Phi_y = E\{\mathbf{y}\mathbf{y}^H\}$  is the  $2M \times 2M$  coherence matrix of the received signal. The coherence matrices  $\Phi_x$  and  $\Phi_y$  are assumed Hermitian positive semi-definite. Equation (5) is a quadratic function of the coefficient vectors  $\mathbf{w}_L$  and  $\mathbf{w}_R$ . Due to its strict convexity, the minimum of  $J_{WF}(\mathbf{w})$  is found in closed form by equating its partial derivatives to zero with respect to the coefficients as:

$$\mathbf{w} = \Theta_y^{-1} \Theta_x \mathbf{q}, \quad (7)$$

The filtered signals at the output of the hearing aids are given by  $z_L = \mathbf{w}_L^H \mathbf{y}$  and  $z_R = \mathbf{w}_R^H \mathbf{y}$ .

#### A. MWF-IC

In [26], the MWF-IC was proposed for noise reduction while preserving the dispersive characteristic of diffuse sound fields. This was achieved by penalizing the MWF cost function by the mismatch between input and output IC of the undesired component. Its cost function was defined as

$$J_{IC}(\mathbf{w}) = J_{WF}(\mathbf{w}) + \alpha J_{IC}^V(\mathbf{w}), \quad (8)$$

in which  $\alpha$  is a frequency dependent weighting parameter that allows a trade-off between noise reduction and IC preservation. The additional penalty term in (8) is given by

$$J_{IC}^V(\mathbf{w}) = |IC_{out}^V(\mathbf{w}) - IC_{in}^V|^2, \quad (9)$$

in which the IC of the noise component, respectively at the output and input, is defined as

$$IC_{out}^V(\mathbf{w}) = \frac{\mathbf{w}^H \Theta_{v_2} \mathbf{w}}{\sqrt{(\mathbf{w}^H \Theta_{v_1} \mathbf{w})(\mathbf{w}^H \Theta_{v_4} \mathbf{w})}}, \quad (10)$$

in which

$$\Theta_{v_1} = \begin{bmatrix} \Phi_v & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \Theta_{v_2} = \begin{bmatrix} \mathbf{0} & \Phi_v \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \Theta_{v_4} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Phi_v \end{bmatrix}; \quad (11)$$

$\Phi_v = E\{\mathbf{v}\mathbf{v}^H\}$  is the coherence matrix of the noise component (resulting in  $\mathbf{w}^H \Theta_{v_2} \mathbf{w} = \mathbf{w}_L^H \Phi_v \mathbf{w}_R$ ,  $\mathbf{w}^H \Theta_{v_1} \mathbf{w} = \mathbf{w}_L^H \Phi_v \mathbf{w}_L$ , and  $\mathbf{w}^H \Theta_{v_4} \mathbf{w} = \mathbf{w}_R^H \Phi_v \mathbf{w}_R$ ); and

$$IC_{in}^V = \frac{\mathbf{q}^T \Theta_{v_2} \mathbf{q}}{\sqrt{(\mathbf{q}^T \Theta_{v_1} \mathbf{q})(\mathbf{q}^T \Theta_{v_4} \mathbf{q})}}. \quad (12)$$

In [27] it was demonstrated that the MWF-IC can be also used to control the ITD of a point noise source. The MWF-IC presents superior performance as compared to both the MWF-ITD [23] and the binaural MWF in preserving the spatial localization of the residual noise under anechoic conditions. However, in reverberant environments (low coherent signals), the  $|IC_{out}^V|$  is as small as  $|IC_{in}^V|$ , and ITD may lose reliability [11] [17].

### IV. PROPOSED METHOD

In free-field situations, sounds received in our ears are perceived as coming from a direction that coincides with the actual physical location of the sound source. However, in reverberant acoustic scenarios, the sound that reaches the ears is composed of the original signal that propagates through the direct path and its multiple reflections that occur due to the

existence of reflective surfaces. This superposition of sounds, which emanates from different positions, results in ITDs and ILDs that, for a significant part of the time, may not correspond to the original direction of the source [11].

It was previously mentioned that IC is determinant for the reliability of the main acoustic binaural cues, such as ILD and ITD [11] [17]. When the received signal is degraded by reflections in a room, the magnitude of the coherence tends to decrease ( $< 1.0$ ) [12]. As stated before, ITD tends to lose its meaning for small coherence values, since there are no waveforms with common characteristics that can be compared to calculate the ITD.

Motivated by the contributions presented in [11] and [12], about the importance of large coherence for the correct localization of acoustic sources in reverberant acoustic scenarios, and assuming an acoustic scenario comprised of one source of interest and a single-point noise source, we propose a new method for noise reduction in binaural hearing aids with spatial preservation of the residual noise. This method, called MWF-CB, is a variation of the MWF with artificial intensification of the coherence of the processed noise. Its purpose is to reinforce the ability to locate an undesired point noise source by filtering-out low coherent components inherent to reverberant environments, which may be associated with low reliability binaural cues.

The MWF-CB cost function is given by

$$J_{CB}(\mathbf{w}) = J_{WF}(\mathbf{w}) + \alpha J_{CB}^V(\mathbf{w}), \quad (13)$$

in which

$$J_{CB}^V(\mathbf{w}) = |IC_{out}^V(\mathbf{w})|^{-1}, \quad (14)$$

and  $\alpha$  is a frequency dependent weighting parameter that emphasizes the importance of  $J_{CB}^V(\mathbf{w})$  with relation to  $J_{WF}(\mathbf{w})$ . The minimum value of the auxiliary term  $J_{CB}^V(\mathbf{w})$  occurs when the magnitude of the IC of the residual noise is unitary (since, by definition,  $0 \leq |IC_{out}^V(\mathbf{w})| \leq 1$ ). Differently from the MWF-IC, the MWF-CB is independent of the IC at the input, which reduces computational effort, since it does not need to be evaluated.

Using (5), (10), and (14) in (13) results in

$$J_{CB}(\mathbf{w}) = \mathbf{q}^T \Theta_x \mathbf{q} - \mathbf{q}^T \Theta_x \mathbf{w} - \mathbf{w}^H \Theta_x \mathbf{q} + \mathbf{w}^H \Theta_y \mathbf{w} + \alpha (\mathbf{w}^H \Theta_{v_1} \mathbf{w})^{1/2} (\mathbf{w}^H \Theta_{v_4} \mathbf{w})^{1/2} |\mathbf{w}^H \Theta_{v_2} \mathbf{w}|^{-1}. \quad (15)$$

As  $|IC_{out}^V(\mathbf{w})|^{-1} \geq 1$ , from (10) we can write

$$(\mathbf{w}^H \Theta_{v_1} \mathbf{w})^{1/2} (\mathbf{w}^H \Theta_{v_4} \mathbf{w})^{1/2} \geq |\mathbf{w}^H \Theta_{v_2} \mathbf{w}|. \quad (16)$$

Assuming that, for accurate localization, it is required a large output IC, i. e.  $|IC_{out}^V(\mathbf{w})| \cong 1$ , we have

$$0 \leq (\mathbf{w}^H \Theta_{v_1} \mathbf{w})^{1/2} (\mathbf{w}^H \Theta_{v_4} \mathbf{w})^{1/2} - |\mathbf{w}^H \Theta_{v_2} \mathbf{w}| \ll 1. \quad (17)$$

As a result, the proposed method may be also interpreted as a constrained optimization problem associated to the minimization of the binaural Wiener filter cost function, given by

$$\begin{aligned} & \min_{\mathbf{w}} J_{WF}(\mathbf{w}) \\ & \text{s.t. } (\mathbf{w}^H \Theta_{v_1} \mathbf{w})(\mathbf{w}^H \Theta_{v_4} \mathbf{w}) - (\mathbf{w}^H \Theta_{v_2} \mathbf{w})(\mathbf{w}^H \Theta_{v_2}^H \mathbf{w}) \leq \varepsilon \end{aligned}, \quad (18)$$

in which  $\varepsilon$  is a small positive and real-valued constant, corresponding to the purpose of  $\alpha$  in (15), that establishes the range of the desired output IC. Equation (18) indicates that the

resulting weight vector of (15) is a suboptimal solution of (5).

## V. EXPERIMENTAL SETUP

Simulations were performed with head-related impulse responses (HRIRs) obtained from a multichannel binaural database [32], which are used to emulate signals received by binaural hearing aids. In this database, signals from six microphones were acquired from two behind-the-ear hearing aids ( $M = 3$ ) placed at the ears of a manikin with the shape of a human head and torso. The HRIRs are employed to artificially filter speech and noise to emulate signals coming from different directions of arrival (azimuth, distance and elevation) under different reverberant conditions. In the following results, HRIRs from two reverberant acoustic scenarios with different acoustic characteristics were employed: *Cafeteria* ( $T_{60} \cong 1250$  ms)<sup>3</sup> and *Office I* ( $T_{60} \cong 300$  ms).

Speech signals were selected from [33], in which 10 different sentences spoken by different speakers (5 male and 5 female) were considered. Each sentence has about 3 seconds. The desired acoustic source was simulated as being in front of the dummy head ( $\theta_S = 0^\circ$  azimuth) in all experiments, since it is the most frequent situation during conversation [34].

The interferent signal was obtained from the International Collegium for Rehabilitative Audiology (ICRA) database [35], which was specially designed for hearing aid testing and psychoacoustic assessments. The *track 7* noise was employed, which emulates noise from six simultaneous speakers. This noise occupies the same frequency band as the speech signal.

The speech and noise single-point source condition was assessed for 8 different acoustic scenes:  $S_0N_{-90}$ ,  $S_0N_{-30}$ ,  $S_0N_{52}$  and  $S_0N_{90}$ <sup>4</sup> for *Cafeteria*; and  $S_0N_{-60}$ ,  $S_0N_{-30}$ ,  $S_0N_{30}$  and  $S_0N_{60}$  for *Office I*. Negative azimuths correspond to the left-hand side of the sagittal plane of the manikin, while positive azimuths correspond to the right-hand side. Hence, the ICRA noise was convolved with 8 HRIRs, creating 8 distinct acoustic conditions:  $\theta_N \in \{-90^\circ, -30^\circ, 52^\circ, 90^\circ\}$  in *Cafeteria*, and  $\theta_N \in \{-60^\circ, -30^\circ, 30^\circ, 60^\circ\}$  in *Office I*. Different epochs of *track 7* were considered for obtaining different noise realizations.

For  $S_0N_{-60}$ ,  $S_0N_{-30}$ ,  $S_0N_{30}$  and  $S_0N_{60}$  in *Office I*, both speech and noise sources were at 100 cm from the manikin, with null elevation. For the *Cafeteria* scenario, elevation angle is  $0^\circ$ , while distances from the manikin, for  $S_0N_{-90}$ ,  $S_0N_{-30}$ ,  $S_0N_{52}$ , and  $S_0N_{90}$  are, respectively,  $\{162, 102, 162, 102\}$  cm for the speech source, and  $\{102, 117.5, 129, 162\}$  cm for the interferent source.

The sampling frequency was  $f_s = 16$  kHz, and the input signals were transformed to the frequency domain by an  $N = 256$  bin STFT, with an analysis window of 128 samples, zero padding, and 50% of overlap. The transformed signals in the STFT domain were reconstructed to the time-domain by the weighted overlap-and-add method [36].

Coherence matrices  $\Phi_y$  and  $\Phi_v$  were computed *a priori*, directly from the contaminated signals by simulating an ideal

<sup>3</sup>  $T_{60}$  refers to the reverberation time, which is the time it takes for a sound to decay by 60 dB.

<sup>4</sup>  $S_{\theta_S}N_{\theta_N}$  means the speech source (S) is placed at  $\theta_S$  degrees of azimuth and the noise source (N) is at  $\theta_N$  degrees of azimuth with respect to the head midline.

voice activity detector (VAD). Speech correlation matrices were calculated as  $\Phi_x = \Phi_y - \Phi_v$  [27]. Errors due to the effect of using a real VAD are not considered in this work.

The SNR of the contaminated signal was defined at the ear closest to the noise source, which is called *worse ear*. According to [34], the most common listening situations in which adults with hearing losses are exposed present SNRs from 2 dB to 14 dB. Therefore, experiments were performed within and beyond this range, using input SNR  $\in \{0, 5, 14, 25\}$  dB.

The performance of the proposed MWF-CB method is assessed and compared to both the conventional binaural MWF [19] and the MWF-IC [26].

Since there is no closed-form solution for both MWF-IC and MWF-CB, numerical techniques must be employed. The optimal filters were obtained by applying a quasi-Newton optimization method [37]. The weighting factor  $\alpha$  was kept fixed for all bins and for both methods.

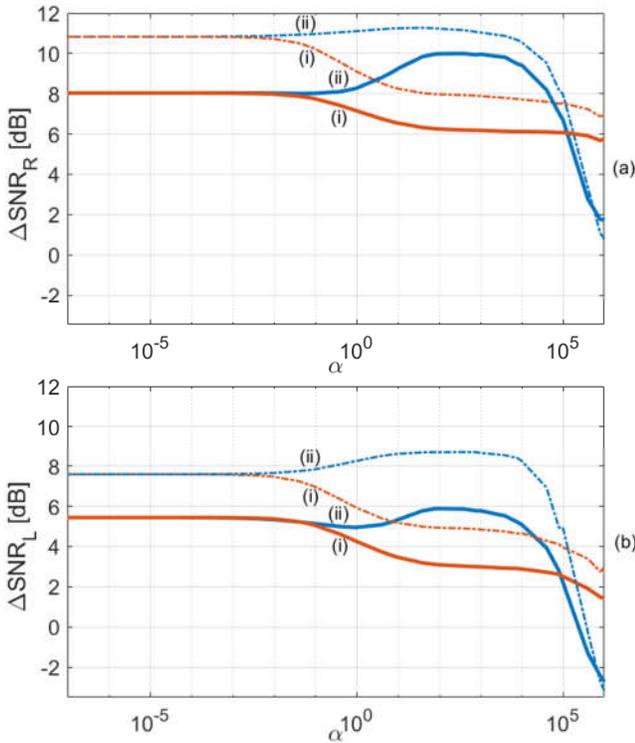


Fig. 2.  $\Delta$ SNR for input SNR = 5 dB as a function of  $\alpha$ : (a) *worse ear*, and (b) *better ear*. (i) MWF-IC (red), and (ii) MWF-CB (blue), for  $S_0N_{52}$  (continuous line) and  $S_0N_{90}$  (dash-dotted line).

A. Objective Measures

Seven objective measures were calculated for assessing the performance of the analyzed methods. They are: (i) signal-to-noise ratio (SNR), which measures noise reduction; (ii) wideband perceptual evaluation of speech quality (PESQ) [38], which measures the overall quality of the enhanced speech signal; (iii) cepstrum distance (CD) [38], which is based on the discrepancy between target and reference signals in the cepstral domain; (iv) short-time objective intelligibility (STOI) [39], which measures the intelligibility of the processed signals; (v) magnitude of the IC of the noise component ( $|JC_{out}^v|$ ), which measures the coherence between

left and right processed signals; (vi) interaural time difference (ITD) [25] calculated up to 1.5 kHz [27]; and (vii) interaural level difference (ILD) [28].

In a first analysis, these objective measures are presented as a function of the weighting parameter ( $\alpha$ ). Following, the MWF-CB and MWF-IC techniques were compared by independently setting their weighting parameters, under some predefined conditions. Three set point conditions were considered: (a) two in which the averaged output SNR in the *worse ear* is the same as the conventional MWF (this approach is similar to the comparison method adopted in [27]); and (b) the one in which the averaged output SNR in the *worse ear* is the highest. The averaged output SNR in the *worse ear* was estimated from computational simulations through 10 runs for each one of the 8 different acoustic scenes described in Section V.

Some results are presented under the output-input difference form ( $\Delta Measure = Measure_{output} - Measure_{input}$ ), in a way that a positive value corresponds to an increase of the objective measure at the output as compared to the input.

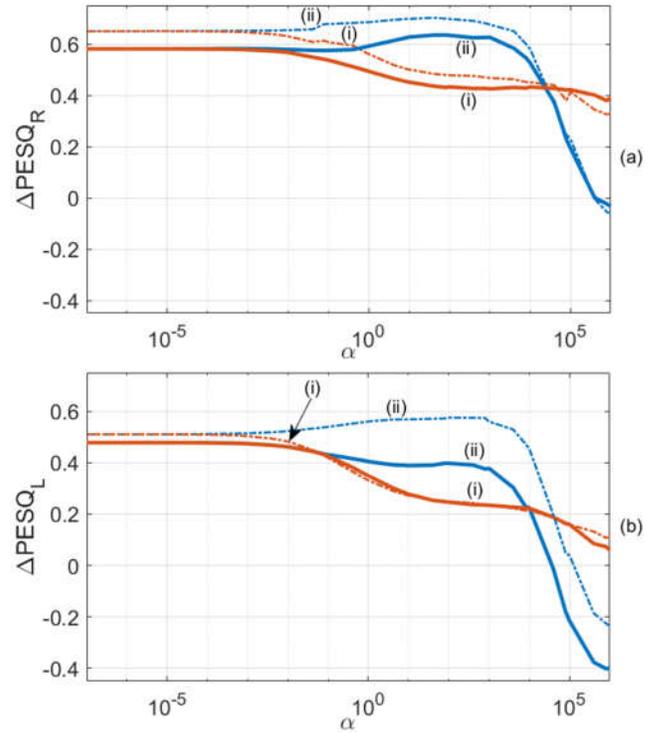


Fig. 3.  $\Delta$ PESQ for input SNR = 5 dB as a function of  $\alpha$ : (a) *worse ear*, and (b) *better ear*. (i) MWF-IC (red), and (ii) MWF-CB (blue), for  $S_0N_{52}$  (continuous line) and  $S_0N_{90}$  (dash-dotted line).

B. Psychoacoustic Experiments

Experiments with volunteers were conducted to evaluate the psychoacoustic aspects of the signals processed by the MWF, MWF-IC, and MWF-CB techniques. The purpose of this experiment is to verify and compare the perceived azimuth of both speech and noise sources. The lateralization experiments were conducted using headphones connected into a personal computer.

The selected group of 10 volunteers comprised 7 males and 3 females, aging between 26 and 59 years (mean of 37.6 and

standard deviation of 13.4 years). No previous complaints regarding hearing losses were declared<sup>5</sup>.

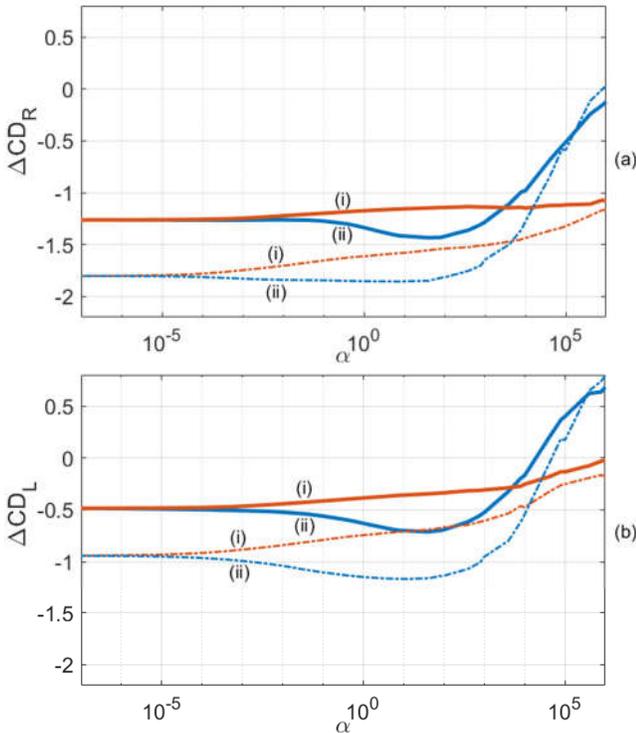


Fig. 4.  $\Delta CD$  for input SNR = 5 dB as a function of  $\alpha$ : (a) *worse ear*, and (b) *better ear*. (i) MWF-IC (red), and (ii) MWF-CB (blue), for  $S_0N_{52}$  (continuous line) and  $S_0N_{90}$  (dash-dotted line).

The experiments consist of three different phases [27]: (a) *learning*, (b) *training*, and (c) *testing*. In the *learning* phase, the volunteers listened to pure ICRA noise filtered by HRIRs related to 13 different azimuths  $\{-90^\circ, -75^\circ, -60^\circ, -45^\circ, -30^\circ, -15^\circ, 0^\circ, 15^\circ, 30^\circ, 45^\circ, 60^\circ, 75^\circ, 90^\circ\}$ , without any additional processing. Visual information (related to the true source azimuth) was synchronized with each audio and presented on the computer screen to guide the volunteers. This phase was employed for volunteers getting familiar with the lateralization task. In order of getting results from highly reliable volunteers only, a verification of this ability was conducted in the *training* phase, in which the volunteers were asked to identify the azimuth of 7 (from the initial set of 13) audios associated to the  $\{-90^\circ, -60^\circ, -30^\circ, 0^\circ, 30^\circ, 60^\circ, 90^\circ\}$  azimuths. These unprocessed audios were presented in random order, with no previous knowledge about the correct azimuths. In this eliminatory phase, volunteers who had perform hemisphere inversions, i.e., lateralization errors related to the left-right sides, were considered not reliable and were excluded from the psychoacoustic experiment. The next phase was the *testing*, in which the remaining volunteers were asked to evaluate a set of 108 audios, including randomly presented speeches and noises composed by unprocessed signals and processed by the MWF, MWF-IC and MWF-CB. A protractor with 13 radio buttons, ranging from  $-90^\circ$  to  $90^\circ$  in steps of  $15^\circ$  (similarly to the method used in [27]) was presented on the

computer screen. In this final stage, the volunteers were asked to identify the perceived azimuth of the selected noises and speeches.

Results from two acoustic scenarios are presented:  $S_0N_{52}$  and  $S_0N_{90}$  from *Cafeteria*. The set of the 108 assessed audios was composed of 50 speech sentences and 50 epochs of noise. The assessed audios were equally divided in five groups: (i) unprocessed signals; and signals filtered by the optimum coefficients obtained with the (ii) MWF, (iii) MWF-IC for  $\alpha = 3 \times 10^{-3}$ , (iv) MWF-CB for  $\alpha = 400$ , and (v) MWF-CB for  $\alpha = 2.5 \times 10^4$ . The remaining 8 audios were related to negative azimuths and were not accounted in the results. These signals were randomly presented for avoiding bias in positive azimuths.

All the experiments with volunteers were approved by the Ethics Committee in Human Research, under certificate 23214619.7.0000.0121 CEP-UFSC.

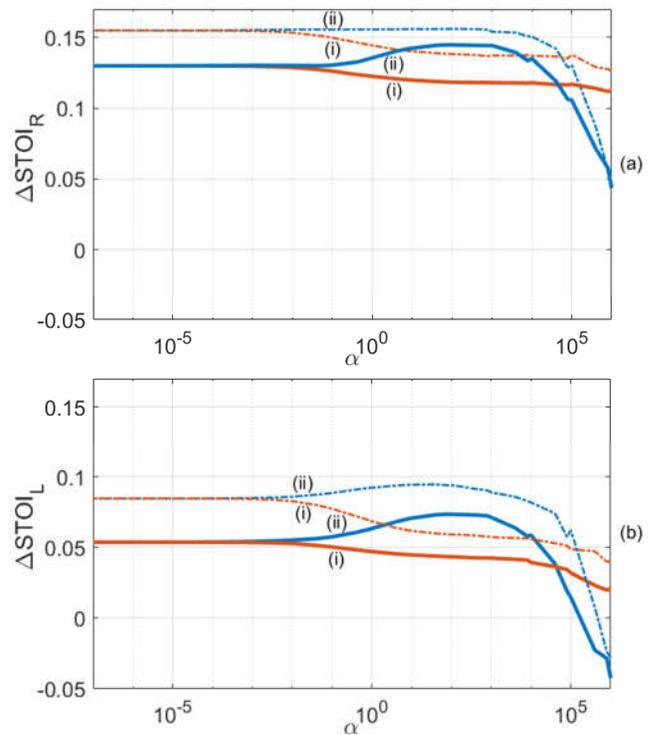


Fig. 5.  $\Delta STOI$  for input SNR = 5 dB as a function of  $\alpha$ : (a) *worse ear*, and (b) *better ear*. (i) MWF-IC (red), and (ii) MWF-CB (blue), for  $S_0N_{52}$  (continuous line) and  $S_0N_{90}$  (dash-dotted line).

## VI. RESULTS

This section presents computational simulations using objective criteria described in Section V.A, followed by results from the psychoacoustic experiments described in Section V.B.

### A. Results from Objective Criteria

Simulations were performed for input SNR  $\in \{0, 5, 14, 25\}$  dB. Since they present approximately the same main characteristics, only results for input SNR = 5 dB are presented, which is the most common input SNR for speech-in-babble noise (such as the ICRA *track 7* noise) [43].

Fig. 2 to Fig. 5 refer to the  $S_0N_{52}$  and  $S_0N_{90}$  (*Cafeteria*)

<sup>5</sup> This procedure is aligned with several earlier works (see [5], [27], [40], [41], and [42]).

scenarios, for both the *worse ear* (R - right ear) and the *better ear* (L - left ear). These results were averaged by 10 runs with different male and female speakers and noise epochs from the *track 7* noise.

Fig. 2 shows the input-output SNR variation ( $\Delta\text{SNR} = \text{SNR}_{\text{output}} - \text{SNR}_{\text{input}}$ ) as a function of the weighting factor  $\alpha$  for both MWF-IC and MWF-CB methods. The plateaus in the extreme left side correspond to the  $\Delta\text{SNR}$  provided by the MWF technique ( $\alpha \rightarrow 0$ ). Differently from the MWF-IC, in which, after the plateau, the  $\Delta\text{SNR}$  decreases with the increase of the weighting factor, the MWF-CB shows a range in which the  $\Delta\text{SNR}$  increases, overcoming the noise-reduction performance of the conventional MWF.

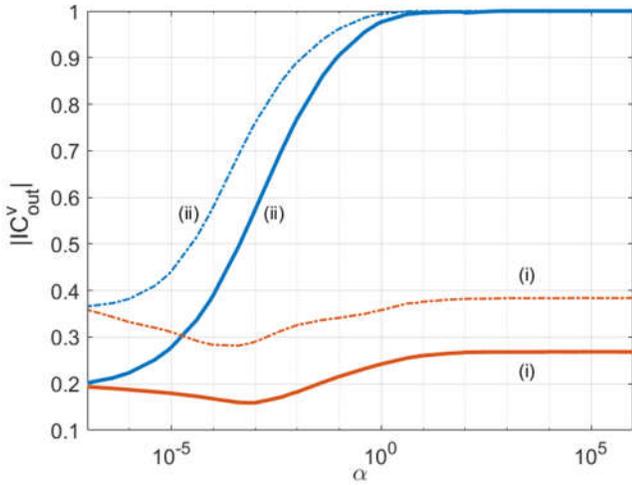


Fig. 6. Magnitude of the IC of the processed noise: (i) MWF-IC (red), and (ii) MWF-CB (blue), for  $S_0N_{52}$  (continuous line) and  $S_0N_{90}$  (dash-dotted line).

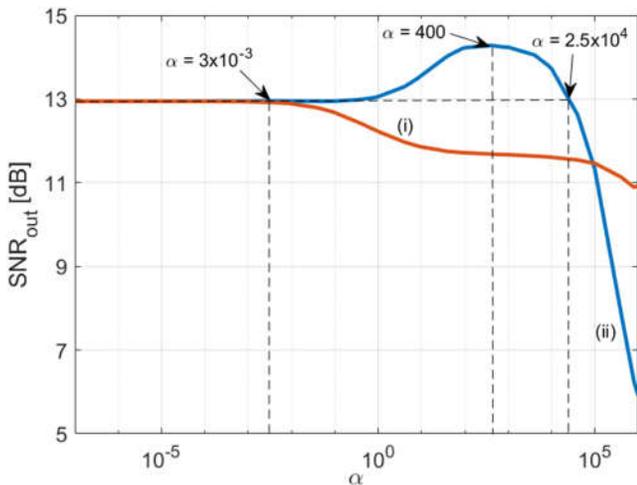


Fig. 7. Averaged output SNR for all acoustic scenarios in the *worse ear*:  $S_0N_{-60}$ ,  $S_0N_{-30}$ ,  $S_0N_{30}$  and  $S_0N_{60}$  for *Office I*; and  $S_0N_{-90}$ ,  $S_0N_{-30}$ ,  $S_0N_{52}$  and  $S_0N_{90}$  for *Cafeteria*. (i) MWF-IC (red), and (ii) MWF-CB (blue). Input SNR is 5 dB. The dash-dotted black lines indicate each weighting factor related to the set point.

Fig. 3 to Fig. 5 show  $\Delta\text{PESQ}$ ,  $\Delta\text{CD}$  and  $\Delta\text{STOI}$  scores. It can be verified a similar behavior to the  $\Delta\text{SNR}$ . Note that negative  $\Delta\text{CD}$  shown in Fig. 4 indicate an enhancement of the processed signal, since lower CD scores are considered better

(meaning lower distance from the target signal in the cepstral domain).

TABLE I  
WEIGHTING FACTORS ( $\alpha$ ) AND ASSOCIATED SNR

	MWF	MWF-IC	MWF-CB		
$\alpha$	0	$3 \times 10^{-3}$	$3 \times 10^{-3}$	400	$2.5 \times 10^4$
$\text{SNR}_{\text{out}}$	12.95	12.95	12.95	14.28	12.95

TABLE II  
OBJECTIVE MEASURES FOR  $S_0N_{52}$  (*CAFETERIA*)

	Raw	MWF	MWF-IC	MWF-CB		
				$3 \times 10^{-3}$	400	$2.5 \times 10^4$
$\text{SNR}_R$	5.02	13.05	13.03	13.23	15.01	13.65
$\text{SNR}_L$	9.08	14.52	14.49	14.47	14.95	13.4
$\text{PESQ}_R$	1.30	1.89	1.86	1.88	1.92	1.74
$\text{PESQ}_L$	1.64	2.14	2.13	2.11	2.03	1.73
$\text{CD}_R$	4.97	3.71	3.73	3.71	3.61	4.18
$\text{CD}_L$	4.23	3.74	3.77	3.72	3.61	4.28
$\text{STOI}_R$	0.74	0.87	0.86	0.87	0.88	0.86
$\text{STOI}_L$	0.81	0.86	0.86	0.86	0.88	0.86
$ IC_{\text{out}}^v $	0.27	0.19	0.17	0.67	1.00	1.00

TABLE III  
OBJECTIVE MEASURES FOR  $S_0N_{90}$  (*CAFETERIA*)

	Raw	MWF	MWF-IC	MWF-CB		
				$3 \times 10^{-3}$	400	$2.5 \times 10^4$
$\text{SNR}_R$	4.99	15.81	15.75	15.84	16.14	14.91
$\text{SNR}_L$	8.26	15.85	15.8	15.89	16.96	15.61
$\text{PESQ}_R$	1.26	1.91	1.90	1.91	1.95	1.70
$\text{PESQ}_L$	1.40	1.91	1.90	1.92	1.97	1.70
$\text{CD}_R$	5.11	3.31	3.36	3.27	3.36	4.11
$\text{CD}_L$	4.42	3.48	3.56	3.41	3.35	4.19
$\text{STOI}_R$	0.78	0.93	0.93	0.94	0.94	0.92
$\text{STOI}_L$	0.84	0.92	0.92	0.93	0.93	0.91
$ IC_{\text{out}}^v $	0.38	0.36	0.31	0.83	1.00	1.00

TABLE IV  
NOISE AND SPEECH  $\Delta\text{ITD}$  AND  $\Delta\text{ILD}$  FOR  $S_0N_{52}$  AND  $S_0N_{90}$  IN *CAFETERIA*

	$\alpha$	NOISE		SPEECH		
		$\Delta\text{ITD}$	$\Delta\text{ILD}$	$\Delta\text{ITD}$	$\Delta\text{ILD}$	
$S_0N_{52}$	MWF		0.36	6.93	0.02	0.25
	MWF-IC		0.33	6.94	0.02	0.26
	MWF-CB	$3 \times 10^{-3}$	0.37	7.18	0.02	1.31
		400	0.54	7.37	0.02	0.26
$S_0N_{90}$	MWF-CB	$2.5 \times 10^4$	0.38	6.69	0.22	3.56
		MWF		0.47	7.10	0.01
	MWF-IC		0.45	7.10	0.01	0.22
	MWF-CB	$3 \times 10^{-3}$	0.48	6.71	0.01	0.20
400		0.48	5.64	0.03	1.14	
	$2.5 \times 10^4$	0.37	5.36	0.17	3.20	

The  $\Delta\text{SNR}$ ,  $\Delta\text{PESQ}$ ,  $\Delta\text{CD}$  and  $\Delta\text{STOI}$  scores presented in Fig. 2 to Fig. 5 show a quite insensitive behavior of the proposed method for a wide range of weighting factors. This

indicates, under certain limiting values of  $\alpha$ , strong robustness with respect to acoustic comfort, speech quality and intelligibility. Further results for  $\Delta$ SNR,  $\Delta$ PESQ,  $\Delta$ CD, and  $\Delta$ STOI for *Office I* and for  $S_0N_{-30}$  and  $S_0N_{-90}$  in *Cafeteria* were omitted for brevity, since they show similar behavior.

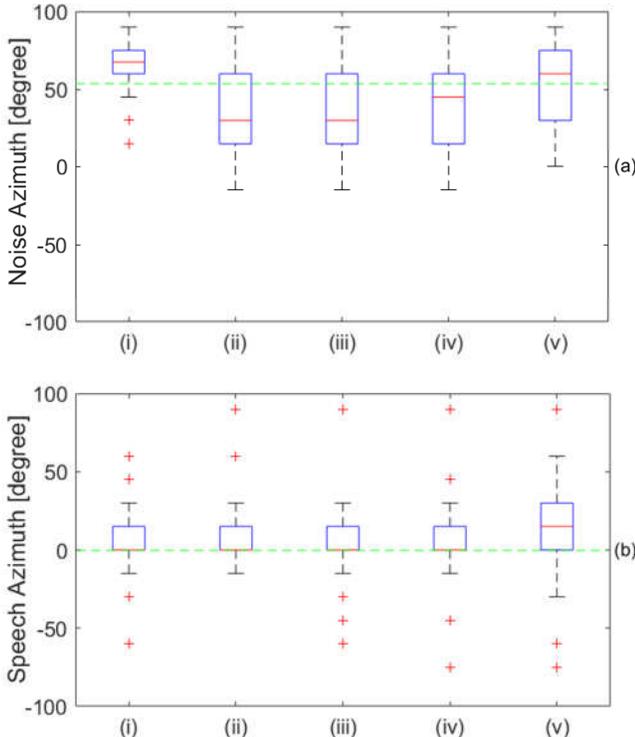


Fig. 8. Boxplots showing how the volunteers perceived the azimuth of (a) noise and (b) speech in the psychoacoustic experiments for  $S_0N_{52}$  in *Cafeteria*. The perceived azimuths are shown for: (i) unprocessed signal; (ii) processed by the MWF; (iii) processed by the MWF-IC for  $\alpha = 3 \times 10^{-3}$ ; (iv) processed by the MWF-CB for  $\alpha = 400$ ; and (v) processed by the MWF-CB for  $\alpha = 2.5 \times 10^4$ . The true azimuth is shown as a dashed green line.

Fig. 6 shows the magnitude of the IC of the processed noise ( $|IC_{out}^V|$ ) as a function of  $\alpha$ . As predicted by the MWF-CB theory, as  $\alpha$  increases, more emphasis is given to the proposed auxiliary term, resulting in  $|IC_{out}^V| \rightarrow 1$ . On the other hand, for the MWF-IC, we have  $|IC_{out}^V| \rightarrow |IC_{in}^V|$ , which demonstrates that the maximum value for  $|IC_{out}^V|$  is limited by the  $|IC_{in}^V|$ , which may be small under reverberant conditions.

The next results present a comparison among the MWF, MWF-IC designed with the largest weighting factor that keeps the original MWF noise reduction capability, and the MWF-CB for three specific set points defined from Fig. 7. In Fig. 7 is shown the global averaged output SNR obtained for all acoustic scenes (4 in *Office I* and 4 in *Cafeteria*), at the *worse ear*, with input SNR = 5 dB, and 80 runs (10 runs per acoustic scene with different male and female speakers and noise epochs). Note that the MWF-CB has a wider range for the weighting factor as compared to the MWF-IC, allowing not only to increase  $\alpha$  without losing noise reduction performance, but even increasing the output SNR.

Table I summarizes the selected  $\alpha$  parameters obtained from results in Fig. 7. Tables II and III show objective scores (SNR, PESQ, CD, STOI and  $|IC_{out}^V|$ ) for, respectively,  $S_0N_{52}$  and  $S_0N_{90}$  obtained for the MWF, MWF-IC and MWF-CB

methods using the  $\alpha$  parameters summarized in Table I. The first (raw) column show scores related to unprocessed noisy-speech. Table IV shows the noise and speech  $\Delta$ ITD and  $\Delta$ ILD for the same scenarios, which suggest a controllable trade-off between noise and speech spatial preservation.

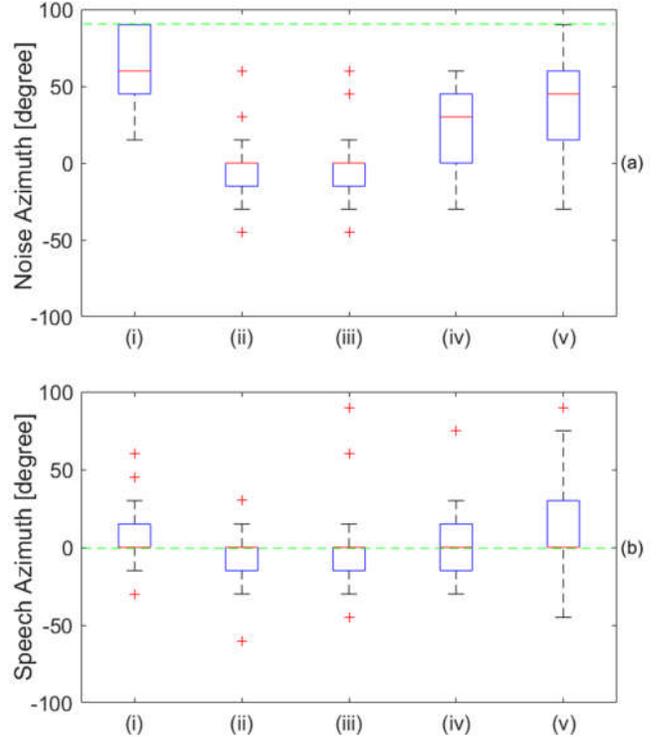


Fig. 9. Boxplots showing how the volunteers perceived the azimuth of (a) noise and (b) speech in the psychoacoustic experiments for  $S_0N_{90}$  in *Cafeteria*. The perceived azimuths are shown for: (i) unprocessed signal; (ii) processed by the MWF; (iii) processed by the MWF-IC for  $\alpha = 3 \times 10^{-3}$ ; (iv) processed by the MWF-CB for  $\alpha = 400$ ; and (v) processed by the MWF-CB for  $\alpha = 2.5 \times 10^4$ . The true azimuth is shown as a dashed green line.

**B. Results from Psychoacoustic Experiments**

Psychoacoustic experiments were carried out according to section V.B. Audio files containing unprocessed signals, and processed by the MWF ( $\alpha = 0$ ), the MWF-IC for  $\alpha = 3 \times 10^{-3}$ , as well as the MWF-CB for  $\alpha = 400$  and  $\alpha = 2.5 \times 10^4$  (see Table I) were assessed. Results are presented in the form of box-and-whisker diagrams [44].

Originally, there were 11 volunteers, but only the results obtained from the 10 most skilled were employed, since one volunteer performed hemisphere inversions during the *training* phase, and was excluded from the experiment.

Fig. 8 shows perceived azimuth results provided by the volunteers for both residual noise (Fig. 8a) and processed speech (Fig. 8b) for the  $S_0N_{52}$  scenario. Similarly, Fig. 9 shows results for  $S_0N_{90}$ . Each boxplot comprises 50 samples. The dashed green line refers to the true source azimuth.

To verify the statistical significance of the data presented in Fig. 8 and Fig. 9, the Shapiro-Wilk (SW) method was applied to test the hypothesis of gaussianity of the distributions ( $H_0$ : all 5 distributions analyzed are Gaussian, at the level of significance  $p > 0.05$ ). As the null hypothesis ( $H_0$ ) was rejected, square root and logarithm transformations were applied. Again,  $H_0$  was rejected. Thus, the Friedman test was

applied ( $H_0$ : all 5 distributions are identical, at the level of significance  $p > 0.05$ ), and the null hypothesis was rejected. Then, the Dunn-Bonferroni post-hoc was applied ( $H_0$ : the pair has the same distribution, at a level of significance  $p > 0.05$ ) to verify which pairs were different from each other. Finally, it was concluded that the pairs summarized in Table V have the same distribution.

Note that in Fig. 8(a)(i) and Fig. 9(a)(i), related to unprocessed noise, there are no hemisphere inversions. This indicates that volunteers provide reasonable accurate estimations for the noise azimuth. The MWF-CB, for both weighting factors, showed a more accurate localization performance as compared to the MWF and MWF-IC.

The perceived azimuth of the speech source is shown in Fig. 8(b) and Fig. 9(b). Medians indicate speech perceived azimuths at  $0^\circ$  for all cases, except for the MWF-CB with the higher  $\alpha$  ( $2.5 \times 10^4$ ) in the  $S_0N_{52}$  scenario, which resulted in  $15^\circ$ .

Additional experiments with the  $S_{60}N_0$ ,  $S_{-60}N_0$ ,  $S_{60}N_{-60}$ ,  $S_{-60}N_{60}$  for the *Office I*, and  $S_{52}N_0$ ,  $S_{-30}N_0$ ,  $S_{52}N_{-52}$ ,  $S_{-30}N_{30}$  for the *Cafeteria* scenarios were also performed with similar results.

TABLE V  
STATISTICAL SIGNIFICANCE FOR THE PSYCHOACOUSTIC EXPERIMENTS ( $P > 0.05$ ): (I) UNPROCESSED SIGNAL; (II) PROCESSED BY THE MWF; (III) PROCESSED BY THE MWF-IC FOR  $\alpha = 3 \times 10^{-3}$ ; (IV) PROCESSED BY THE MWF-CB FOR  $\alpha = 400$ ; AND (V) PROCESSED BY THE MWF-CB FOR  $\alpha = 2.5 \times 10^4$

$S_0N_{52}$	METHODS	P-VALUE	$S_0N_{90}$	METHODS	P-VALUE
NOISE	(I)-(V)	0.463	NOISE	(I)-(V)	0.095
	(II)-(III)	1.0		(II)-(III)	1.0
	(II)-(IV)	1.0		(IV)-(V)	0.072
	(III)-(IV)	1.0			
SPEECH	(I)-(II)	1.0	SPEECH	(I)-(II)	0.114
	(I)-(III)	1.0		(I)-(III)	0.065
	(I)-(IV)	1.0		(I)-(IV)	1.0
	(I)-(V)	0.059		(I)-(V)	1.0
	(II)-(III)	1.0		(II)-(III)	1.0
	(II)-(IV)	1.0		(II)-(IV)	0.82
	(III)-(IV)	1.0		(III)-(IV)	0.537
				(IV)-(V)	0.766

## VII. DISCUSSION

The relevance of the proposed cost function is based on the fact that the conventional binaural MWF has a noticeable noise reduction capacity for hearing aid applications, but results in distortion of the binaural cues of the residual noise. Consequently, the perception of the original azimuth of the noise source is changed towards to the speech azimuth.

The MWF-IC aims to overcome this drawback, allowing a controlled trade-off between spatial preservation of the noise source and noise reduction performance. However, reverberant conditions in the acoustic scenario may decrease the noise IC at the input microphones. Hence, preserving poorly coherent components, i.e., the input IC and associated ITD, may not be sufficient to allow an accurate perception of the true azimuth of the noise source.

The noise reduction approach proposed in (13) was

motivated by the results presented in [11] and [12] about the importance of the IC for the correct localization of acoustic sources, as well as its influence over the reliability of ITD and ILD binaural cues. The auxiliary cost function presented in (14) aims to regularize the optimization problem, intensifying the output IC of the noise, despite the input IC condition. This, of course, limits its applicability to the single-point noise-source case.

Objective criteria were applied for performance evaluation of the proposed technique, which show a strong robustness of the MWF-CB with regard to acoustic comfort, speech quality and intelligibility, for a wide range of weighting factors (see Fig. 2 to Fig. 5).

A very important observation is the fact that the proposed MWF-CB is capable of providing a higher noise reduction performance as compared to both binaural MWF and MWF-IC. This happens due to the artificial emphasis of the IC. Results obtained indicate that the proposed method may provide an increase of up to 2 dB SNR for certain scenarios.

The maximum value of  $|IC_{out}^v|$  for the MWF-IC is limited by  $|IC_{in}^v|$  (see Fig. 6), which may be a small value under reverberant conditions. Note, however, that  $|IC_{out}^v|$  for the MWF-CB tends to 1.0 as  $\alpha$  increases, whatever the IC input conditions.

Three distinct set points for the MWF-CB were compared to the MWF-IC and MWF (see Table I). Using these set points, the intelligibility, acoustic comfort and quality objective criteria for both MWF-IC and MWF-CB show scores similar to the MWF (see Tables II and III). On the other hand, a bigger value of  $|IC_{out}^v|$  can be obtained with the proposed method (see Fig. 6).

Psychoacoustic results for both  $S_0N_{52}$  and  $S_0N_{90}$  show that the MWF-CB provides the best preservation of the true noise azimuth (see Fig. 8(a) and Fig. 9(a)). From the medians, it can be noticed that the speech azimuth was accurately identified in all situations (see Fig. 8(b) and Fig. 9(b)). The only exception was for the  $S_0N_{52}$  scenario, in which the MWF-CB for  $\alpha = 2.5 \times 10^4$  presented an error of  $15^\circ$ . This distortion of the perceived azimuth of the speech is in agreement with  $\Delta ITD$  and  $\Delta ILD$  objective measures presented in Table IV.

Consequences of this side-effect on the perception of the speech azimuth observed for bigger  $\alpha$  may be alleviated considering that, in the most common listening situations, the speaker is in front of the hearing aid user ( $\theta_S = 0^\circ$ ) [34] [45] (to allow lip-reading). In this situation, a small distortion in the perceived azimuth is not a major problem.

Despite the presented objective criteria for several values of  $\alpha$ , further listening experiments indicate that extremely higher values of the weighting factor ( $\alpha \gg 2.5 \times 10^4$ ) decrease noise reduction, overall quality and intelligibility performance, as well as distort the perception of the true azimuth of the desired speech source. This occurs because the prevalence of  $J_{CB}^v$  over  $J_{WF}$  in (13).

In addition to these reported performance features, the proposed technique also shows better computational properties. Massive simulations performed in a desktop personal computer with an Intel® Core i7-4790 processor, running at 3.90 GHz, and Matlab®, indicated that the numerical process for obtaining the optimal filters ( $w_L$  and  $w_R$ )

for the MWF-CB is 3.9 times faster than for the MWF-IC.

## VIII. CONCLUSION

This work presented a new MWF-based technique for binaural hearing aids that provides perceptually relevant spatial enhancement for low-coherent acoustic signals, such as those received from a single-point noise-source in reverberant environments.

Objective criteria and psychoacoustic experiments indicate an intensification of the original noise interaural coherence, as well as higher output signal-to-noise ratio as compared to the conventional MWF and MWF-IC methods. As a counterpart, distortions on speech binaural cues may occur in extreme conditions. Such trade-off is controlled by a weighting parameter, which should be carefully designed.

Additional simulation results provide strong evidence that the proposed method has better optimization characteristics as compared to a previous developed competing technique (MWF-IC), reducing the processing time required for calculating the optimum filters.

## REFERENCES

- [1] H. Levitt, "Noise reduction in hearing aids: a review," *J. Rehab. Res. Dev.*, vol. 38, no. 1, pp. 111-121, 2001.
- [2] S. Kochkin, "MarkeTrak VIII: Consumer satisfaction with hearing aids is slowly increasing," *Hear. J.*, vol. 63, no. 1, pp. 19-32, 2010. doi: 10.1097/01.HJ.0000366912.40173.76
- [3] V. Hamacher, *et al.*, "Binaural signal processing in hearing aids: Technologies and algorithms," in *Advances in Digital Speech Transmission*, New York: Wiley, 2008, pp. 401-429.
- [4] P. A. Naylor and N. D. Gaubitch, *Speech Dereverberation*. London: Springer, 2010.
- [5] J. Thiemann, *et al.*, "Speech enhancement for multimicrophone binaural hearing aids aiming to preserve the spatial auditory scene," *EURASIP J. Adv. Signal Process.*, vol. 2016, no. 12, pp. 1-11, 2016. doi: 10.1186/s13634-016-0314-6
- [6] M. Jeub *et al.*, "Model-based dereverberation preserving binaural cues," *IEEE Trans. Audio Speech Lang. Process.*, vol. 18, no. 7, pp. 1732-1745, 2010. doi: 10.1109/TASL.2010.2052156
- [7] T. Bogaert *et al.*, "Speech enhancement with multichannel Wiener filter techniques in multimicrophone binaural hearing aids," *J. Acoust. Soc. America*, vol. 125, no. 1, pp. 360-371, 2009. doi: 10.1121/1.3023069
- [8] J. Blauert, *Spatial Hearing: the psychophysics of human sound localization*, MIT press, 1997.
- [9] M. Jeub, M. Dorbecker, and P. Vary, "A semi-analytical model for the binaural coherence of noise fields," *IEEE Signal Process. Lett.*, vol. 18, no. 3, pp. 197-200, 2011. doi: 10.1109/LSP.2011.2108284
- [10] T. Bogaert *et al.*, "Horizontal localization with bilateral hearing aids: without is better than with," *J. Acoust. Soc. Am.*, vol. 119, no. 1, pp. 515-526, 2006. doi: 10.1121/1.2139653
- [11] C. Faller and J. Merimaa, "Source localization in complex listening situations: Selection of binaural cues based on interaural coherence," *J. Acoust. Soc. Am.*, vol. 116, no. 5, pp. 3075-3089, 2004. doi: 10.1121/1.1791872
- [12] B. Rakerd and W. Hartmann, "Localization of sound in rooms. V. Binaural coherence and human sensitivity to interaural time differences in noise," *J. Acoust. Soc. Am.*, vol. 128, no. 5, pp. 3052-3063, 2010. doi: 10.1121/1.3493447
- [13] J. Werner, M. Costa, and F. Itturriet, "Noise reduction method based on the MWF and in the interaural coherence for binaural hearing aids," (in Portuguese) in *Proc. Braz. Simp. Telecommun. Signal Process (SBrt)*, 2019, pp. 1-5. doi: 10.14209/SBRT.2019.1570539306
- [14] W. Hartmann, "Localization of sound in rooms," *J. Acoust. Soc. Am.*, vol. 74, pp. 1380-1391, 1983.
- [15] B. Rakerd and W. Hartmann, "Localization of sound in rooms. II. The effects of a single reflecting surface," *J. Acoust. Soc. Am.*, vol. 78, pp. 524-533, 1985.
- [16] M. Bodden, "Modeling human sound source localization and the cocktail-party-effect," *Acta Acustica*, vol. 1, no. 122, pp. 43-55, 1993.
- [17] M. Dietz, S. Ewert, and V. Hohmann, "Auditory model based direction estimation of concurrent speakers from binaural signals," *Speech Commun.*, vol. 53, pp. 592-605, 2011. doi: 10.1016/j.specom.2010.05.006
- [18] E. Hadad, S. Doclo, and S. Gannot, "The binaural LCMV beamformer and its performance analysis," *IEEE Trans. Audio Speech Lang. Process.*, vol. 24, no. 3, pp. 543-558, 2016. doi: 10.1109/TASLP.2016.2514496
- [19] S. Doclo *et al.*, "Acoustic beamforming for hearing aid applications," in *Handbook on Array Processing and Sensor Networks*. New York: Wiley, 2010, pp. 269-302. doi: 10.1002/9780470487068
- [20] S. Braun *et al.*, "Evaluation and comparison of late reverberation power spectral density estimators," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 26, no. 6, pp. 1052-1067, 2018. doi: 10.1109/TASLP.2018.2804172
- [21] B. Cornelis *et al.*, "Theoretical analysis of binaural multimicrophone noise reduction techniques," *IEEE Trans. Audio Speech Lang. Process.*, vol. 18, no. 2, pp. 342-355, 2010. doi: 10.1109/TASL.2009.2028374
- [22] D. Marquardt, V. Hohmann, and S. Doclo, "Interaural coherence preservation in multi-channel Wiener filtering-based noise reduction for binaural hearing aids," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 23, no. 12, pp. 2162-2176, 2015. doi: 10.1109/TASLP.2015.2471096
- [23] S. Doclo *et al.*, "Extension of the multi-channel Wiener filter with ITD cues for noise reduction in binaural hearing aids," in *IEEE Workshop App. Signal Process. Audio Acoust.*, 2005, pp. 70-73. doi: 10.1109/ASPAA.2005.1540171
- [24] S. Doclo *et al.*, "Extension of the multi-channel Wiener filter with localization cues for noise reduction in binaural hearing aids," in *IEEE Int. Workshop Acoust. Signal Enhanc. (IWAENC)*, 2005, pp. 221-224.
- [25] T. Bogaert *et al.*, "Binaural cue preservation for hearing aids using an interaural transfer function multichannel Wiener filter," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2007, pp. 565-568. doi: 10.1109/ICASSP.2007.366975
- [26] D. Marquardt, V. Hohmann, and S. Doclo, "Coherence preservation in multichannel Wiener filtering based noise reduction for binaural hearing aids," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2013, pp. 8648-8652. doi: 10.1109/ICASSP.2013.6639354
- [27] F. Itturriet and M. Costa, "Perceptually relevant preservation of interaural time differences in binaural hearing aids," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 27, no. 4, pp. 753-764, 2019. doi: 10.1109/TASLP.2019.2895973
- [28] M. Costa and P. Naylor, "ILD preservation in the multichannel Wiener filter for binaural hearing aid applications," in *European Signal Process. Conf. (EUSIPCO)*, 2014, pp. 636-640.
- [29] A. Westermann, J. Buchholz, and T. Dau, "Binaural dereverberation based on interaural coherence histograms," *J. Acoust. Soc. Am.*, vol. 133, no. 5, pp. 2767-2777, 2013. doi: 10.1121/1.4799007
- [30] K. Simmer, S. Fischer, and A. Wasiljeff, "Suppression of coherent and incoherent noise using a microphone array," *Annales des Télécommun.*, vol. 49, no. 7-8, pp. 439-446, 1994.
- [31] I. McCowan and H. Bourlard, "Microphone array post-filter based on noise field coherence," *IEEE Trans. Audio Speech Lang. Process.*, vol. 11, no. 6, pp. 709-716, 2003. doi: 10.1109/TSA.2003.818212
- [32] H. Kayser *et al.*, "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses," *EURASIP J. Adv. Signal Process.*, vol. 2009, pp. 1-10, 2009. doi: 10.1155/2009/298605
- [33] Recommendation ITU-T P.50, Telephone Transmission Quality, Telephone Installations, Local Line Networks: Objective Measuring Apparatus – Artificial Voices, Appendix I: Test Signals, 1998.
- [34] Y. Wu *et al.*, "Characteristics of real-world signal to noise ratios and speech listening situations of older adults with mild to moderate hearing loss," *Ear Hear.*, vol. 39, no. 2, pp. 293-304, 2018. doi: 10.1097/AUD.0000000000000486
- [35] W. Dreschler *et al.*, "ICRA noises: artificial noise signals with speech-

like spectral and temporal properties for hearing instrument assessment,” *Audiol.*, vol. 40, no. 3, pp. 148-157, 2001. doi: 10.3109/00206090109073110

- [36] R. Crochiere, “A weighted overlap-add method of short-time Fourier analysis/synthesis,” *IEEE Trans. Acoust. Speech Signal Process.*, vol. 28, no. 1, pp. 99-102, 1980. doi: 10.1109/TASSP.1980.1163353
- [37] J. Arora, *Introduction to Optimum Design*, 2<sup>nd</sup> ed. New York: Elsevier, 2004.
- [38] Y. Hu and P. Loizou, “Evaluation of objective quality measures for speech enhancement,” *IEEE Trans. Audio Speech Lang. Process.*, vol. 16, pp. 229-238, 2008. doi: 10.1109/TASL.2007.911054
- [39] C. Taal *et al.*, “A short-time objective intelligibility measure for time-frequency weighted noisy speech,” in *IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2010, pp. 4214-4217. doi: 10.1109/ICASSP.2010.5495701
- [40] D. Carmo and M. Costa, “Online approximation of the multichannel Wiener filter with preservation of interaural level difference for binaural hearing-aids,” *Comput. Biol. Med.*, vol. 95, pp.188-197, 2018. doi: <https://doi.org/10.1016/j.combiomed.2018.02.017>
- [41] R. Borges and M. Costa, “A feed forward adaptive canceller to reduce the occlusion effect in hearing aids,” *Comput. Biol. Med.*, vol. 79, pp. 266-75, 2016. doi: <https://doi.org/10.1016/j.combiomed.2016.10.016>
- [42] M. Costa, “A complementary low-cost method for broadband noise reduction in hearing aids for medium to high SNR levels,” *Comput. Biol. Med.*, vol. 46, pp. 29-41, 2014. doi: <https://doi.org/10.1016/j.combiomed.2013.12.009>
- [43] K. Smeds, F. Wolters, and M. Rung, “Estimation of signal-to-noise ratios in realistic sound scenarios,” *J. Am. Acad. Audiol.*, vol. 26, no. 2, pp.183-196, 2015. doi: 10.3766/jaaa.26.2.7
- [44] D. Shannon, “Box-and-whisker plots with the SAS,” *Pharm. Stat.*, vol. 2, pp. 291-295, 2003.
- [45] B. Walden *et al.*, “Predicting hearing aid microphone preference in everyday listening,” *J. Am. Acad. Audiol.*, vol. 15, no. 5, 365-396, 2004. doi: 10.3766/jaaa.15.5.4



**Johnny Werner** received the B.E.E. degree in 2008 and the M.Sc. degree in power electronics in 2012, both from Fundação Universidade Regional de Blumenau, Blumenau, Brazil. He is currently working toward the Ph.D. degree with Universidade Federal de Santa Catarina, Florianópolis, Brazil. Since 2013, he has been with the Electrical Academic Department, Universidade Tecnológica Federal do Paraná (UTFPR), Pato Branco, Brazil. His present research interests include speech dereverberation, noise reduction and binaural cue preservation in hearing aids.



**Márcio Holsbach Costa** received the B.E.E. degree from Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil, in 1991; the M.Sc. degree in biomedical engineering from Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil, in 1994; and the Dr. degree in electrical engineering from Universidade Federal de Santa Catarina (UFSC), Florianópolis, Brazil, in 2001. From 1994 to 2004 he was with the Department of Electrical Engineering, Biomedical Engineering Group, Universidade Católica de Pelotas (UCPel). Since 2004 he is with the Department of Electrical and Electronic Engineering at Universidade Federal de Santa Catarina. In 2013, he was a Visiting Researcher with the Communications and Signal Processing Research Group, Imperial College London. His present research interests are in biomedical signal processing, hearing aids, cochlear implants, and active noise and vibration control.