Evaluation of Signal-Dependent Partial Noise Estimation Algorithms for Binaural Hearing Aids

Jonas Klug[1], *Daniel Marquardt*[1,2], *Nico Gößling*[1], *Simon Doclo*[1]

[1]Department of Medical Physics and Acoustics and Cluster of Excellence Hearing4all, University of Oldenburg, Germany
[2]Starkey Hearing Technologies, USA
Email: {jonas.klug,daniel.marquardt,nico.goessling,simon.doclo}@uni-oldenburg.de

Abstract

Besides noise reduction an important objective of a binaural speech enhancement algorithm is the preservation of the binaural cues of both the target speaker as well as the undesired noise sources. Although the binaural minimum variance distortionless response (BMVDR) beamformer results in a good noise reduction performance and preserves the binaural cues of the target speaker, it distorts the binaural cues of the background noise, such that the target speaker and the background noise are perceived as coming from the same direction. Aiming at also preserving the binaural cues of the background noise, the BMVDR beamformer with partial noise estimation (BMVDR-N) was proposed, where a parameter allows to trade-off between noise reduction and binaural cue preservation of the background noise. In this paper, we propose a signal-dependent method to determine this trade-off parameter based on the coherence between the noisy input signals and the output signals of the BMVDR beamformer. Simulation results and subjective listening tests for a realistic acoustic scenario with diffuse cafeteria noise show that the proposed signal-dependent trade-off parameter for the BMVDR-N beamformer significantly improves the spatial quality compared to the BMVDR beamformer, while achieving the same speech intelligibility.

1 Introduction

Noise reduction algorithms for hearing aids are crucial to improve speech quality and speech intelligibility in background noise [1]. Besides reducing noise, preserving the binaural cues of all sound sources is an important objective of a binaural noise reduction algorithm in order to ensure that the listener's impression of the acoustic scene is not distorted. The binaural cues are not only important for spatial awareness, but also have an impact on speech intelligibility due to so-called binaural unmasking [2, 3].
Binaural hearing aids, consisting of one or more microphones on each side of the head of the listener, are able to exploit not only spectral but also spatial information. The binaural minimum variance distortionless response (BMVDR) beamformer aims at minimizing the power of the output signals while providing an undistorted response for the target speaker [4, 5]. Due to this constraint the binaural cues of the target speaker are not affected. On the other hand, it has been shown in [4, 5] that the binaural cues of all other sound sources, i.e. background noise and interfering speakers, are not preserved, such that all sound sources are perceived as coming from the direction of the target speaker. Considering the importance of the binaural cues for speech intelligibility (SI) and spatial awareness, several binaural algorithms have been proposed that aim at combining binaural cue preservation of all sound sources and noise reduction [6–18].
Aiming at preserving the binaural cues of the noise component, in [4] the BMVDR beamformer with partial noise estimation (BMVDR-N) was proposed. In the BMVDR-N beamformer a parameter determines the amount of mixing between the output signals of the BMVDR beamformer (maximum noise reduction but no preservation of the binaural cues) and the input signals (no noise reduction but perfect preservation of the binaural noise cues). Hence this parameter allows to trade-off between noise reduction and preservation of the binaural noise cues., i.e. between SI and spatial awareness, several methods have been proposed to determine this trade-off parameter. Whereas in [4] a broadband trade-off was used in [15, 16] a frequency-dependent trade-off parameter was proposed that is based on the interaural coherence (IC) discrimination abilities of the human auditory system in a diffuse noise field. Alternatively, it has been proposed in [17, 18] to either use the output signals of the BMVDR beamformer or a scaled version of the noisy input signals depending on whether the target speaker or the noise is dominant in a time-frequency bin.
In this paper, we propose a novel signal-dependent method to determine the frequency-dependent trade-off parameter of the BMVDR-N beamformer, where this parameter depends on the magnitude squared coherence (MSC) between the noisy microphone signals and the output signals of the BMVDR beamformer. For a realistic acoustic scenario with a single target speaker and diffuse cafeteria noise, we compare the performance of the BMVDR-N beamformer using the proposed MSC-based trade-off parameter and using the IC-based trade-of parameter from [15] with the BMVDR beamformer. We evaluate the considered algorithms both using objective measures (hybrid intelligibility-weighted SNR improvement, binaural spatial quality [19]) and subjective listening tests. The objective and subjective results show that the proposed method improves the spatial quality compared to the BMVDR beamformer, while achieving similar speech intelligibility. Compared to the method from [16], the proposed method achieves a better performance, both in terms of speech intelligibility and spatial quality.

This work was supported in part by the Cluster of Excellence 1077 Hearing4all, funded by the German Research Foundation (DFG), and by the joint Lower Saxony-Israeli Project ATHENA.

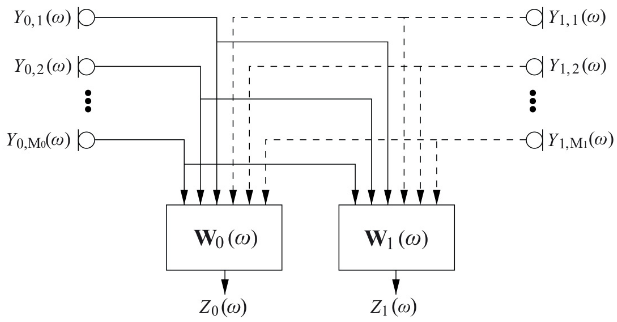


Figure 1: General binaural hearing aid setup.

2 Notation

Consider the binaural hearing aid setup in Fig.1 with M_0 microphones on the left hearing aid and M_1 microphones on the right hearing aid. The m-th microphone signal on the left hearing aid is defined in the frequency domain as

$$Y_{0,m}(\omega) = X_{0,m}(\omega) + N_{0,m}(\omega), \tag{1}$$

with ω the normalized radian frequency, $X_{0,m}$ the desired speech component and $N_{0,m}$ the background noise component. For conciseness we will omit the frequency variable ω in the remainder of the paper. We define the M-dimensional stacked signal vector $\boldsymbol{Y}$ with $M = M_0 + M_1$, as

$$\boldsymbol{Y} = [Y_{0,1} ... Y_{0,M_0}\ Y_{1,1} ... Y_{1,M_1}]^T. \tag{2}$$

This vector can be written as

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{N}, \tag{3}$$

where the vectors $\boldsymbol{X}$ and $\boldsymbol{N}$ are defined similarly as $\boldsymbol{Y}$ in (2). Considering an acoustical scenario with a single target speaker S_x, the speech vector $\boldsymbol{X}$ can be written as

$$\boldsymbol{X} = S_\mathrm{x}\boldsymbol{A}, \tag{4}$$

with $\boldsymbol{A}$ denoting the acoustic transfer functions (ATFs) between the target speaker and the microphones. Without loss of generality, we will use the first microphone signal of the left and the right hearing aid as the reference microphone signals, i.e.

$$Y_0 = \mathbf{e}_0^T\boldsymbol{Y}, \quad Y_1 = \mathbf{e}_1^T\boldsymbol{Y}, \tag{5}$$

where $\mathbf{e}_0$ and $\mathbf{e}_1$ are M-dimensional selector vectors with one element equal to 1 and all other elements equal to 0, i.e. $\mathbf{e}_0(1)=1$, $\mathbf{e}_1(M_0+1)=1$. The relative transfer function (RTF) vectors $\bar{\boldsymbol{A}}_0$ and $\bar{\boldsymbol{A}}_1$ for the left and the right hearing aid are defined by relating the ATF vector $\boldsymbol{A}$ to the ATF of the reference microphone on the left and the right hearing aid respectively, i.e.

$$\bar{\boldsymbol{A}}_0 = \frac{\boldsymbol{A}}{A_0}, \quad \bar{\boldsymbol{A}}_1 = \frac{\boldsymbol{A}}{A_1}. \tag{6}$$

The $M \times M$-dimensional covariance matrices of the speech component and the noise component are defined as

$$\boldsymbol{R}_\mathrm{x} = \mathscr{E}\left\{\boldsymbol{X}\boldsymbol{X}^H\right\}, \quad \boldsymbol{R}_\mathrm{n} = \mathscr{E}\left\{\boldsymbol{N}\boldsymbol{N}^H\right\}, \tag{7}$$

where $\mathscr{E}$ denotes the expectation operator. The output signals on the left and the right hearing aid are filtered versions of the input signals, i.e.

$$Z_0 = \boldsymbol{W}_0^H\boldsymbol{Y} = Z_{\mathrm{x}0} + Z_{\mathrm{n}0} = \boldsymbol{W}_0^H\boldsymbol{X} + \boldsymbol{W}_0^H\boldsymbol{N}, \tag{8}$$

$$Z_1 = \boldsymbol{W}_1^H\boldsymbol{Y} = Z_{\mathrm{x}1} + Z_{\mathrm{n}1} = \boldsymbol{W}_1^H\boldsymbol{X} + \boldsymbol{W}_1^H\boldsymbol{N}, \tag{9}$$

where $\boldsymbol{W}_0$ and $\boldsymbol{W}_1$ denote the M-dimensional complex-valued filter vectors for the left and right hearing aid, and $Z_{\mathrm{x}0}$, $Z_{\mathrm{n}0}$, $Z_{\mathrm{x}1}$, $Z_{\mathrm{n}1}$ denote the speech and noise components in the left and the right hearing aid, respectively.

3 Binaural Noise Reduction

This section briefly reviews two state-of-the-art binaural noise reduction algorithms, i.e. the BMVDR beamformer [4, 5] and the BMVDR-N beamformer [4, 16].

3.1 Binaural MVDR Beamformer

The BMVDR beamformer aims at minimizing the power spectral density (PSD) of the noise component in the output signals on both hearing aids while preserving the speech component in the reference microphone signals, i.e.

$$\min_{\boldsymbol{W}_0} \mathscr{E}\left\{|\boldsymbol{W}_0^H\boldsymbol{N}|^2\right\} \text{ subject to } \boldsymbol{W}_0^H\boldsymbol{A} = A_0, \tag{10}$$

$$\min_{\boldsymbol{W}_1} \mathscr{E}\left\{|\boldsymbol{W}_1^H\boldsymbol{N}|^2\right\} \text{ subject to } \boldsymbol{W}_1^H\boldsymbol{A} = A_1, \tag{11}$$

Using (6) and (7), the resulting filter vectors of the BMVDR beamformer can be written in terms of the RTF vectors as

$$\boldsymbol{W}_{\mathrm{BMVDR},0} = \frac{\boldsymbol{R}_\mathrm{n}^{-1}\bar{\boldsymbol{A}}_0}{\bar{\boldsymbol{A}}_0^H\boldsymbol{R}_\mathrm{n}^{-1}\bar{\boldsymbol{A}}_0}, \; \boldsymbol{W}_{\mathrm{BMVDR},1} = \frac{\boldsymbol{R}_\mathrm{n}^{-1}\bar{\boldsymbol{A}}_1}{\bar{\boldsymbol{A}}_1^H\boldsymbol{R}_\mathrm{n}^{-1}\bar{\boldsymbol{A}}_1}. \tag{12}$$

As shown in [5, 15], the BMVDR beamformer preserves the binaural cues of the target speaker, but distorts the IC of the noise component in such a way that the speech and the residual noise are perceived as directional sources coming from the same direction, which is obviously not desired.

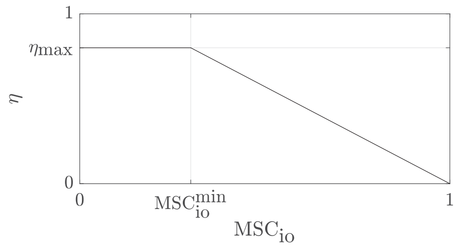


Figure 2: Characteristic curve between the magnitude squared coherence $\mathrm{MSC_{io}}$ and the trade-off parameter η.

3.2 Binaural MVDR Beamformer with Partial Noise Estimation (BMVDR-N)

Aiming at also preserving the binaural cues of the noise component, it has been proposed in [4] to modify the BMVDR beamformer. Instead of aiming at completely suppressing the noise component, the BMVDR-N beamformer aims at preserving a portion of the noise component, i.e.

$$\min_{\boldsymbol{W}_0} \mathscr{E}\left\{|\boldsymbol{W}_0^H\boldsymbol{N} - \eta_0 N_0|^2\right\} \text{ subject to } \boldsymbol{W}_0^H\boldsymbol{A} = A_0, \tag{13}$$

$$\min_{\boldsymbol{W}_1} \mathscr{E}\left\{|\boldsymbol{W}_1^H\boldsymbol{N} - \eta_1 N_1|^2\right\} \text{ subject to } \boldsymbol{W}_1^H\boldsymbol{A} = A_1. \tag{14}$$

where η_0 and η_1 denote the trade-off parameter for the left and the right hearing aid, which are typically equal, i.e. $\eta = \eta_0 = \eta_1$. It has been shown in [4] that the filter vectors of the BMVDR-N beamformer can be written as

$$\boldsymbol{W}_{\text{BMVDR-N},0} = (1-\eta_0)\boldsymbol{W}_{\mathrm{BMVDR},0} + \eta_0\mathbf{e}_0, \tag{15}$$

$$\boldsymbol{W}_{\text{BMVDR-N},1} = (1-\eta_1)\boldsymbol{W}_{\mathrm{BMVDR},1} + \eta_1\mathbf{e}_1. \tag{16}$$

Hence, the output signals of the BMVDR-N beamformer are equal to the sum of the BMVDR beamformer output signals (scaled with 1-η) and the reference microphone signals (scaled with η). This means that for $\eta = 0$ maximum noise reduction is achieved but the binaural cues of the noise component are not preserved.

In [16] a psychoacoustically motivated method has been proposed to compute the (frequency-dependent) trade-off parameter such that a desired IC for the residual noise component is obtained. It should be noticed that this approach is independent of the SNR. Alternatively, it has been proposed in [17, 18] to either use the BMVDR beamformer output signals or a scaled version of the noisy reference microphone signals, depending on whether the speech component or the noise component is dominant in a time-frequency bin. These approaches are motivated by the assumption that in time-frequency bins with a large SNR (where noise is hardly audible) it is more important to keep the maximum noise reduction provided by the BMVDR beamformer than to preserve the binaural cues of the noise component, whereas in time-frequency bins with a low SNR (where noise is audible), it is more important to preserve the binaural cues of the complete acoustic scene.

4 MSC-dependent Trade-Off Parameter

Building upon the SNR-dependent approaches from [17, 18], in this paper we propose to set the trade-off parameters of the BMVDR-N beamformer as a continuous function of the SNR and to set them independently for the left and the right hearing aid. Instead of directly using the SNR, which may not be straightforward to estimate, we propose to use the magnitude squared coherence (MSC) between the noisy reference microphone signals and the output signals of the BMVDR beamformer, i.e.

$$\mathrm{MSC}_{\mathrm{io},0} = \frac{|\mathscr{E}\{Y_0\, Z_0^*\}|^2}{\mathscr{E}\{|Y_0|^2\}\mathscr{E}\{|Z_0|^2\}}, \tag{17}$$

ISBN 978-3-8007-4767-2

$$\mathrm{MSC}_{\mathrm{io},1} = \frac{|\mathscr{E}\{Y_1\, Z_1^*\}|^2}{\mathscr{E}\{|Y_1|^2\}\mathscr{E}\{|Z_1|^2\}}. \tag{18}$$

Large $\mathrm{MSC}_{\mathrm{io}}$ corresponds to a large SNR [11], while a small $\mathrm{MSC}_{\mathrm{io}}$ corresponds to a small SNR. Therefore we propose to use the characteristic curve depicted in Fig. 2 to set the trade-off parameters η_0 and η_1 as a function of $\mathrm{MSC}_{\mathrm{io},0}$ and $\mathrm{MSC}_{\mathrm{io},1}$, respectively. For $\mathrm{MSC}_{\mathrm{io}} \leq \mathrm{MSC}_{\mathrm{io}}^{\mathrm{min}}$, η is equal to η_{max}, whereas for $\mathrm{MSC}_{\mathrm{io}} > \mathrm{MSC}_{\mathrm{io}}^{\mathrm{min}}$, η is linearly decaying with $\mathrm{MSC}_{\mathrm{io}}$. Using such a characteristic curve, a large η is used for time-frequency bins with a small SNR, whereas a small η is used for time-frequency bins with a large SNR.

5 Experimental Results

In this section we compare the performance of the considered binaural beamformers, i.e. the BMVDR beamformer and the BMVDR-N beamformer, either using the IC-based trade-off parameter proposed in [16] (BMVDR-N-IC) or using the proposed MSC-based trade-off parameters discussed in section 4 (BMVDR-N-MSC). In Section 5.1 the used acoustic setup is presented. In section 5.2 the implementation of the beamformers and the used performance measures are discussed. In Section 5.3 the results of the objective performance measures and the subjective listening tests are presented.

5.1 Acoustic scenario

To generate the microphone signals, we used measured impulse responses for a binaural hearing aid setup in a cafeteria ($T_{60} \approx 1250\,\mathrm{ms}$) [20]. Each hearing aid mounted on a dummy head was equipped with 2 microphones, i.e. $M = 4$. The reverberant speech component was generated by convolving clean speech signals from the Oldenburg Sentence Test (OLSA) database [21] with the measured impulse responses. Ambient noise including babble noise, clacking plates and interfering speakers, recorded in the same cafeteria, was used as the noise component [20]. The target speaker was either in front of the dummy head (scenario 1, 0°) or on the left of the dummy head (scenario 2, -35°). The broadband input SNR was set to -5 dB and the sample frequency was equal to 16 kHz.

5.2 Implementation and Performance Measures

All time-domain signals were transformed to the frequency-domain using the short-time Fourier transform (STFT) with frame length $N = 480$ samples and frame shift $P = 240$ samples, e.g., for the reference microphone signal on the left hearing aid

$$Y_0(k,l) = \sum_{n=0}^{N-1} y_0(lP+n)W(n)e^{\frac{-j2\pi kn}{N}}, \tag{19}$$

with k the frequency index, l the frame index and $W(n)$ the analysis window.

Assuming a diffuse noise field, we used the spatial coherence matrix $\mathbf{\Gamma}$ instead of the noise covariance matrix $\boldsymbol{R}_{\mathrm{n}}$ in (12) for the BMVDR, BMVDR-N-IC and the BMVDR-N-MSC beamformers. The (i,j)-th element of the spatial coherence matrix $\mathbf{\Gamma}$ is calculated as

$$\mathbf{\Gamma}(i,j) = \frac{\sum_{s=1}^{S} A_i(\theta_s)A_j^*(\theta_s)}{\sqrt{\sum_{s=1}^{S}|A_i(\theta_s)|^2 \sum_{s=1}^{S}|A_j^*(\theta_s)|^2}}, \tag{20}$$

with $A(\theta_s)$ denoting the anechoic ATF for a source at angle θ_s and $S = 72$ the total number of angles. The ATFs $A(\theta_s)$ and the RTFs of the target speakers $\bar{\boldsymbol{A}}$ in (12) were calculated from the measured (anechoic) head related impulse responses [20], assuming perfect knowledge of the direction of arrival of the target speaker.

The performance of the considered binaural beamformers is evaluated in terms of noise reduction and preservation of the spatial impression of the acoustic scene. The noise reduction performance of beamformers is often evaluated using the global intelligibility-weighted output SNR [22], which is defined, e.g. for the left hearing aid, as

$$\mathrm{iSNR}_0^{\mathrm{out}} = \sum_{k=1}^{K} I(k) 10\log_{10}(\mathrm{SNR}_0^{\mathrm{out}}(k)), \tag{21}$$

with $I(k)$ a weighting function according to [23], taking into account the importance of the different frequency bands to SI, and

$$\mathrm{SNR}_0^{\mathrm{out}}(k) = \frac{\sum_{l=0}^{L-1}|Z_{\mathrm{x}0}(k,l)|^2}{\sum_{l=0}^{L-1}|Z_{\mathrm{n}0}(k,l)|^2}, \tag{22}$$

where L denotes the number of frames. We now suggest two modifications which are important for evaluating binaural MVDR-based beamformers. First, since MVDR-based beamformers using anechoic RTFs typically perform some dereverberation of the target speaker, which is not harmful for SI, using (22) typically underestimates the performance of the BMVDR beamformer and overestimates the performance of the BMVDR-N beamformer in terms of SI improvement. Hence, instead of using the *processed* speech component $Z_{\mathrm{x}0}$, we propose to use the *unprocessed* speech component X_0, i.e.

$$\overline{\mathrm{SNR}}_0^{\mathrm{out}}(k) = \frac{\sum_{l=0}^{L-1}|X_0(k,l)|^2}{\sum_{l=0}^{L-1}|Z_{\mathrm{n}0}(k,l)|^2}. \tag{23}$$

Secondly, since independently calculating iSNR in the left and the right hearing aid does not take better-ear glimpsing [24] into account, we propose to compute a hybrid SNR based on a monaural better-ear glimpsed speech and noise component, i.e

$$\overline{\mathrm{SNR}}_{\mathrm{h}}^{\mathrm{out}}(k) = \frac{\sum_{l=0}^{L-1}|X_{\mathrm{mono}}(k,l)|^2}{\sum_{l=0}^{L-1}|Z_{\mathrm{n,\,mono}}(k,l)|^2}, \tag{24}$$

with

$$X_{\mathrm{mono}}(k,l) = \begin{cases} X_0(k,l) & \text{if } \frac{|X_0(k,l)|^2}{|Z_{\mathrm{n}0}(k,l)|^2} \geq \frac{|X_1(k,l)|^2}{|Z_{\mathrm{n}1}(k,l)|^2}, \\ X_1(k,l) & \text{if } \frac{|X_0(k,l)|^2}{|Z_{\mathrm{n}0}(k,l)|^2} < \frac{|X_1(k,l)|^2}{|Z_{\mathrm{n}1}(k,l)|^2}, \end{cases} \tag{25}$$

$$Z_{\mathrm{n,mono}}(k,l) = \begin{cases} Z_{\mathrm{n}0}(k,l) & \text{if } \frac{|X_0(k,l)|^2}{|Z_{\mathrm{n}0}(k,l)|^2} \geq \frac{|X_1(k,l)|^2}{|Z_{\mathrm{n}1}(k,l)|^2}, \\ Z_{\mathrm{n}1}(k,l) & \text{if } \frac{|X_0(k,l)|^2}{|Z_{\mathrm{n}0}(k,l)|^2} < \frac{|X_1(k,l)|^2}{|Z_{\mathrm{n}1}(k,l)|^2}. \end{cases} \tag{26}$$

Similarly to (21), the global intelligibility-weighted hybrid SNR is defined as

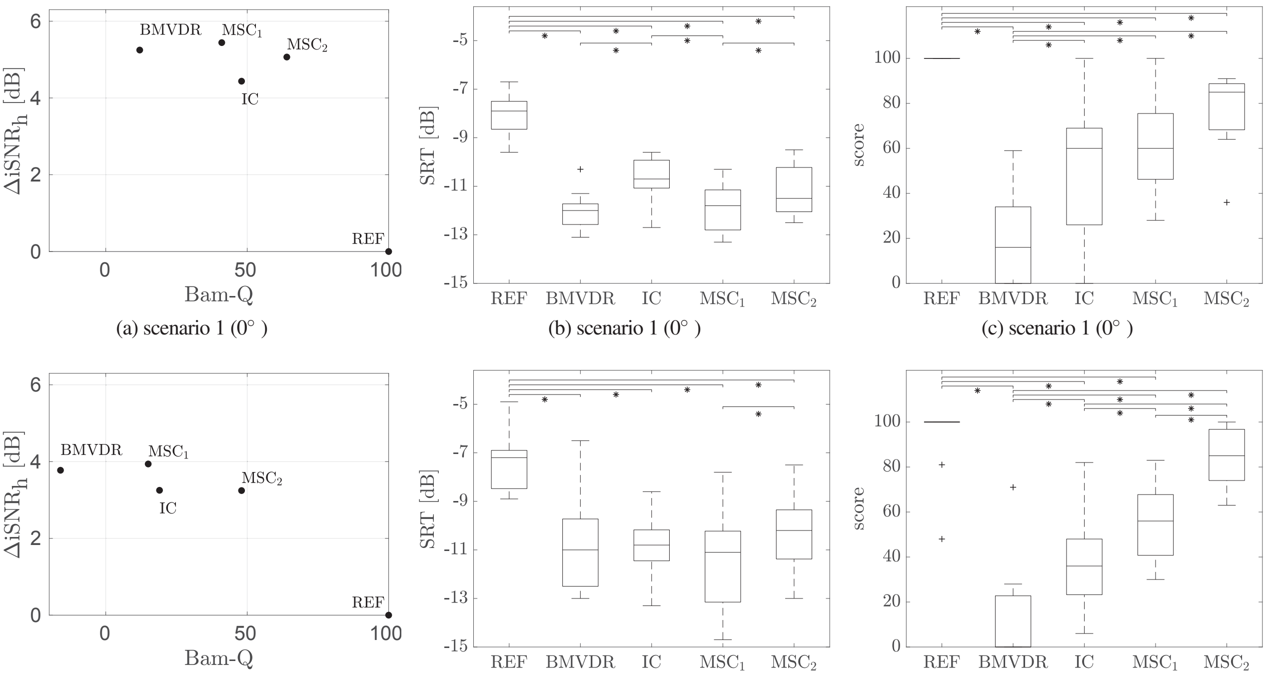
$$\mathrm{iSNR}_{\mathrm{h}}^{\mathrm{out}} = \sum_{k=1}^{K} I(k) 10\log_{10}(\overline{SNR}_h^{\mathrm{out}}(k)). \tag{27}$$

To evaluate the spatial impression of the acoustic scene, we propose to use the psycho-acoustically motivated Bam-Q measure proposed in [19], which is based on the binaural auditory model of [25]. Bam-Q is an intrusive measure, which predicts spatial quality differences between a binaural test signal and a binaural reference signal (in our case the reference microphone signals). A Bam-Q value of 100 corresponds to no difference, 0 to an obvious difference and negative values to even larger difference.

The trade-off parameter η for the BMVDR-N-IC beamformer was computed according to [16]. For the MSC-based trade-off parameters of the BMVDR-N-MSC beamformer, the $\mathrm{MSC}_{\mathrm{io}}$ in (17) and (18) were computed by recursively updating the required PSDs and cross-PSDs with a time constant of 20 ms. Based on multiple simulations, we consider two characteristic curves: BMVDR-N-MSC_1 with $\eta_{\mathrm{max}} = 0.7$ and $\mathrm{MSC}_{\mathrm{io}}^{\mathrm{min}} = 0$, which yielded the best performance in terms of $\mathrm{iSNR}_{\mathrm{h}}^{\mathrm{out}}$ and BMVDR-N-MSC_2 with $\eta_{\mathrm{max}} = 1$ and $\mathrm{MSC}_{\mathrm{io}}^{\mathrm{min}} = 0.1$, which yielded the best performance in terms of Bam-Q while performing at least as well as the MVDR-N-IC in terms of $\mathrm{iSNR}_{\mathrm{h}}^{\mathrm{out}}$.

5.3 Objective and Subjective Results

In this section we compare the performance of the considered binaural beamformers (BMVDR, BMVDR-N-IC, BMVDR-N-MSC_1, BMVDR-N-MSC_2) in terms of speech intelligibility and spatial awareness, based on objective performance measures (Section 5.2) and subjective listening tests.



(a) scenario 1 (0°) (b) scenario 1 (0°) (c) scenario 1 (0°)

(d) scenario 2 (-35°) (e) scenario 2 (-35°) (f) scenario 2 (-35°)

Figure 3: Intelligibility-weighted hybrid SNR improvement ($\Delta iSNR_h$) and spatial quality (Bam-Q) (left), SRT results (middle) and MUSHRA scores (right) for the unprocessed input signals and the considered binaural beamformers for both scenarios. On each box, the central mark is the median value and the edges of the box are the 25% and 75% percentiles. The whiskers indicate the largest/smallest value that is smaller/larger than the 75%/25% percentile plus/minus 1.5 times the interquartile range. The outliers are plotted individually. The stars depict significance.

5.3.1 Objective Evaluation

For the objective evaluation we present the average results for 3 randomly picked OLSA sentences. For both scenarios Figure 3 (left) depicts the performance of the four considered binaural beamformers in terms of the intelligibility-weighted hybrid SNR improvement and the spatial quality difference (Bam-Q) to the reference microphone signals. It can be observed that the BMVDR beamformer significantly improves $iSNR_h$ compared to the unprocessed input signals from the reference microphones, but yields a much lower Bam-Q. Compared to the BMVDR beamformer all binaural beamformers with partial noise estimation improve Bam-Q and lead to a comparable $iSNR_h$, where the proposed BMVDR-N-MSC$_2$ yields the best Bam-Q at approximately the same $iSNR_h$ as the BMVDR-N-IC and the proposed BMVDR-N-MSC$_1$ yields the best $iSNR_h$ at approximately the same Bam-Q as the BMVDR-N-IC. For scenario 2 (-35°), it appears that for all beamformers a lower $iSNR_h$ and a lower Bam-Q is obtained then for scenario 1 (0°).

5.3.2 Subjective listening tests

Speech intelligibility was evaluated using the Oldenburg sentence test [21] and spatial quality was evaluated using the MUltiple Stimuli with Hidden Reference and Anchor (MUSHRA) test [26] with N=11 normal hearing subjects. For the MUSHRA test the reference microphone signals were used as a hidden reference and the BMVDR output signals as the anchor. The statistical significance for the subjective test results was analysed using a repeated measures Analysis of Variance (ANOVA) [27] where the normal distribution of the data was tested with Kolmogorow-Smirnow test [28].
Fig. 3 (middle) depicts the speech reception threshold (SRT) results. For both scenarios, all considered binaural beamformers significantly improve the SRT compared to the unprocessed input signals and there is a significant difference between the BMVDR-N-MSC$_1$ and the BMVDR-N-MSC$_2$ beamformer. Moreover, for scenario 1 there is a significant SRT difference between the BMVDR and the BMVDR-N-IC beamformer and between the BMVDR-N-MSC$_1$ and the BMVDR-N-IC beamformer. In general, it can be observed that these SRT differences can be predicted rather well by the objective $iSNR_h$ measure.

Fig. 3 (right) depicts the MUSHRA scores. First, it can be observed that for both scenarios the BMVDR beamformer - as expected - achieves the lowest scores. Secondly, it can be observed that for both scenarios all binaural beamformers with partial noise estimation significantly improve the MUSHRA score compared to the BMVDR beamformer. For scenario 2, there is a significant difference between both BMVDR-N-MSC beamformers and the BMVDR-N-IC beamformer and between the BMVDR-N-MSC$_1$ and the BMVDR-N-MSC$_2$ beamformer. Moreover, for scenario 2 there is not even a significant difference between the proposed BMVDR-N-MSC$_2$ beamformer and the input signals. In general, it can be observed that the MUSHRA scores can be predicted rather well by the objective Bam-Q measure.

6 Conclusion

In this paper we proposed a signal-dependent method to determine the trade-off parameters of the BMVDR-N beamformer. More in particular, we proposed two characteristic curves to compute the tade-off parameter based on the MSC between the noisy microphone signals and the BMVDR output signals. For a realistic acoustic scenario with diffuse cafeteria noise, objective and subjective results show that both proposed BMVDR-N-MSC beamformers achieve a similar speech intelligibility as the BMVDR beamformer, while significantly improving the binaural spatial quality, where the spatial quality of the BMVDR-N-MSC$_2$ is even almost similar to the input signal.

References

[1] S. Doclo, W. Kellermann, S. Makino, and S. E. Nordholm, “Multichannel signal enhancement algorithms for assisted listening devices: Exploiting spatial diversity using multiple microphones,” *IEEE Signal Processing Magazine*, vol. 32, pp. 18–30, March 2015.

[2] J. Blauert, *Spatial hearing : the psychophysics of human sound localization*. Cambridge, Mass. MIT Press, 1997.

[3] M. L. Hawley, R. Y. Litovsky, and J. F. Culling, "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," *The Journal of the Acoustical Society of America*, vol. 115, no. 2, pp. 833–843, 2004.

[4] B. Cornelis, S. Doclo, T. Van den Bogaert, J. Wouters, and M. Moonen, "Theoretical analysis of binaural multi-microphone noise reduction techniques," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 18, pp. 342–355, Feb. 2010.

[5] S. Doclo, S. Gannot, D. Marquardt, and E. Hadad, "Binaural Speech Processing with Application to Hearing Devices," in *Audio Source Separation and Speech Enhancement*, ch. 18, Wiley, 2018.

[6] T. Lotter and P. Vary, "Dual-channel speech enhancement by superdirective beamforming," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. 1–14, 2006.

[7] R. Aichner, H. Buchner, M. Zourub, and W. Kellermann, "Multi-channel source separation preserving spatial information," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (Honolulu HI, USA), pp. 5–8, Apr. 2007.

[8] A. H. Kamkar-Parsi and M. Bouchard, "Instantaneous Binaural Target PSD Estimation for Hearing Aid Noise Reduction in Complex Acoustic Environments," *IEEE Trans. on Instrumentation and Measurement*, vol. 60, pp. 1141–1154, Apr. 2011.

[9] K. Reindl, Y. Zheng, A. Schwarz, S. Meier, R. Maas, A. Sehr, and W. Kellermann, "A stereophonic acoustic signal extraction scheme for noisy and reverberant environments," *Computer Speech and Language*, vol. 27, no. 3, pp. 726–745, 2013.

[10] R. Baumgärtel, M. Krawczyk-Becker, D. Marquardt, C. Völker, H. Hu, T. Herzke, G. Coleman, K. Adiloğlu, S. M. A. Ernst, T. Gerkmann, S. Doclo, B. Kollmeier, V. Hohmann, and M. Dietz, "Comparing binaural signal processing strategies I: Instrumental evaluation.," *Trends in Hearing*, vol. 19, pp. 1–16, 2015.

[11] H. Asad, M. Bouchard, and H. Kamkar-Parsi, "Perceptually motivated binaural beamforming with cues preservation for hearing aids," *in Proc., IEEE Canadian Conference on Electrical and Computer Engineering*, pp. 909–913, May 2016.

[12] D. Marquardt, E. Hadad, S. Gannot, and S. Doclo, "Theoretical Analysis of Linearly Constrained Multi-channel Wiener Filtering Algorithms for Combined Noise Reduction and Binaural Cue Preservation in Binaural Hearing Aids," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, vol. 23, pp. 2384–2397, Dec. 2015.

[13] E. Hadad, S. Doclo, and S. Gannot, "The Binaural LCMV Beamformer and its Performance Analysis," *IEEE/ACM Trans. on Audio, Speech, and Language Proc.*, vol. 24, no. 3, pp. 543–558, 2016.

[14] A. I. Koutrouvelis, R. C. Hendriks, R. Heusdens, and J. Jensen, "Relaxed Binaural LCMV Beamforming," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, vol. 25, pp. 137–152, Jan. 2017.

[15] 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. on Audio, Speech, and Language Processing*, vol. 23, pp. 2162–2176, Dec. 2015.

[16] D. Marquardt and S. Doclo, "Interaural Coherence Preservation for Binaural Noise Reduction Using Partial Noise Estimation and Spectral Postfiltering," *IEEE/ACM Trans. on Audio, Speech and Language Processing*, vol. 26, no. 7, pp. 1257–1270, 2018.

[17] J. Thiemann, M. Müller, D. Marquardt, S. Doclo, and S. van de Par, "Speech enhancement for multimicrophone binaural hearing aids aiming to preserve the spatial auditory scene," *EURASIP Journal on Advances in Signal Processing*, vol. 2016, p. 12, Feb. 2016.

[18] A. I. Koutrouvelis, J. Jensen, M. Guo, R. C. Hendriks, and R. Heusdens, "Binaural Speech Enhancement with Spatial Cue Preservation Utilising Simultaneous Masking," in *Proc. European Signal Processing Conference (EUSIPCO)*, (Kos, Greece), pp. 628–632, Aug. 2017.

[19] J.-H. Fleßner, S. D. Ewert, B. Kollmeier, and R. Huber, "Quality assessment of multi-channel audio processing schemes based on a binaural auditory model," *in Proc., IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1340–1344, May 2014.

[20] H. Kayser, S. Ewert, J. Anemüller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses," *Eurasip Journal on Advances in Signal Processing*, vol. 2009, p. 10, 2009.

[21] K. Wagener, V. Kuehnel, and B. Kollmeier, "Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests (Development and evaluation of a German sentence test I: Design of the Oldenburg sentence test)," *Zeitschrift für Audiologie / Audiological Acoustics*, vol. 38, p. 4–15, 1999.

[22] J. E. Greenberg, P. M. Peterson, and P. M. Zurek, "Intelligibility-weighted measures of speech-to-interference ratio and speech system performance," *The Journal of the Acoustical Society of America*, vol. 94, no. 5, pp. 3009–3010, 1993.

[23] ANSI S3.5-1997, *Methods for calculation of the speech intelligibility index*. American National Standard (ANSI), 1997.

[24] V. Best, C. R. Mason, G. Kidd, N. Iyer, and D. S. Brungart, "Better-ear glimpsing in hearing-impaired listeners," *The Journal of the Acoustical Society of America*, vol. 137, no. 2, pp. EL213–EL219, 2015.

[25] M. Dietz, S. D. Ewert, and V. Hohmann, "Auditory model based direction estimation of concurrent speakers from binaural signals," *Speech Communication*, vol. 53, pp. 592–605, May 2011.

[26] ITU-R BS.1534-3, *Method for the subjective assessment of intermediate quality level of audio systems*. International Telecommunications Union Recommendation sector(ITU-R), 10 2015.

[27] A. Field, *Discovering statistics using IBM SPSS statistics*. Sage Publications, 2013.

[28] G. Kanji, *100 Statistical Tests*. 100 Statistical Tests, Sage Publications, 2006.