A Class of Pareto Optimal Binaural Beamformers

Elior Hadad ⓘ, Simon Doclo ⓘ, *Senior Member, IEEE*, Sven Nordholm ⓘ, *Senior Member, IEEE*, and Sharon Gannot ⓘ, *Fellow, IEEE*

Abstract—The objective of binaural multi-microphone speech enhancement algorithms can be viewed as a multi-criteria design problem as there are several requirements to be met. The objective is not only to extract the target speaker without distortion, but also to suppress interfering sources (e.g., competing speakers) and ambient background noise, while preserving the auditory impression of the complete acoustic scene. Such a multi-objective problem (MOP) can be solved using a Pareto frontier, which provides a useful trade-off between the different criteria. In this paper, we propose a unified Pareto optimization framework, which is achieved by defining a generalized mean squared error (MSE) cost function, derived from a MOP. The solution to the multi-criteria problem is grounded on a solid mathematical foundation. The MSE cost function consists of a weighted sum of speech distortion (SD), partial interference reduction (IR), and partial noise reduction (NR) terms with scaling parameters that control the amount of IR and NR. The filter minimizing this generalized cost function, denoted Pareto optimal binaural multichannel Wiener filter (Pareto-BMWF), constitutes a generalization of various binaural MWF-based and binaural MVDR-based beamformers. This solution is optimal for any set of parameters. The improved speech enhancement capabilities are experimentally demonstrated using real-signal recordings when estimation errors are present and the binaural cue preservation capabilities are analyzed.

Index Terms—Beamforming, binaural cue preservation, hearing aids, LCMV, multi-microphone noise reduction, MVDR, MWF, Pareto optimization.

I. INTRODUCTION

THE objective of binaural noise reduction algorithms is not only to selectively extract the target speaker and to suppress interfering sources and ambient background noise, but also to preserve the auditory impression for the hearing aid user. This can be achieved by preserving the binaural cues of all sound sources, i.e., the interaural level difference (ILD) and the interaural time difference (ITD) for coherent sources (target and interfering sources) and the interaural coherence (IC) for incoherent sound fields (background noise) [1]. These binaural cues play a major role in spatial perception, i.e., the ability to localize sound sources and to determine the spatial width or

diffuseness of a sound field [2], and are very important for speech intelligibility because of the so-called, binaural unmasking effect [3], [4].

Unlike monaural noise reduction algorithms, binaural noise reduction algorithms need to generate two output signals (i.e., one for each ear), hence typically processing all available microphone signals from both devices by using two different spatial filters [5]–[27]. In [8], [14] the binaural minimum variance distortionless response (MVDR) beamformer was proposed, which extracts the target speech component in both hearing aids without distortion while minimizing the overall noise power. In [8], [11], the binaural multichannel Wiener filtering (MWF) was proposed, which provides a minimum MSE estimate of the target speech component in both hearing aids, providing a trade-off between SD and NR. It was also shown that by setting the trade-off parameter to zero, the obtained binaural MWF is equivalent to the binaural MVDR beamformer.

In [28], it was shown that the (monaural) MWF problem can be viewed as a multi-objective problem (MOP), as two competing requirements need to be met, i.e., the minimization of the residual noise power and the SD power [29]. Hence, only non-inferior solutions can be obtained, which are known as Pareto optimal solutions [30].

It was shown that both the binaural MVDR and the binaural MWF preserve the binaural cues of the target speech source but typically distort the binaural cues of the overall noise (i.e., interfering sources and background noise), since all sources are perceived as arriving from the target direction. To preserve the spatial perception for the hearing aid user, several extensions of these binaural beamformers were proposed. In [7], [24], the binaural MWF with partial noise estimation was proposed, which is aimed to preserve the binaural cues of the overall noise while sacrificing noise reduction. Other extensions proposed in [12]–[17], [21] focus on the interfering sources by adding a (hard) interference reduction constraint to the cost function of the binaural MVDR and the binaural MWF, thereby preserving the binaural cues of the interfering sources.

In this paper, we propose a unified Pareto optimization framework for multi-microphone speech enhancement, which is achieved by defining a MOP that consists of SD, partial IR, and NR objective terms. This solution is optimal for any set of parameters. The solution to the multi-criteria problem is grounded on a solid mathematical foundation. Using the scalarization method, we unify the MOP into one single scalar function, i.e., a generalized cost function that consists of a weighted sum of SD, partial IR, and partial NR MSE cost function terms. The filter that minimizes the generalized cost function, denoted as the Pareto-BMWF, defines a Pareto solution to the problem. The

proposed binaural beamformer constitutes a generalization of several binaural MWF-based and binaural MVDR-based beamformers. We show that several known binaural beamformers are special cases for specific settings of the trade-off parameters such that the filters are part of the respective Pareto solution set.

A well-known procedure in the literature (cf. [8]) states that setting the trade-off parameter (between SD and NR) to zero, the binaural MWF reduces to the binaural MVDR beamformer. We claim that such a procedure is not mathematically founded, since the cost function of the binaural MWF becomes ill-posed. The formulated Pareto-BMWF beamformer circumvents these problems by showing that both the binaural MWF and the binaural MVDR beamformer lie on a Pareto frontier of the same MOP.

Our contribution is five-fold. First, we introduce a new formulation for the binaural noise reduction problem that is based on a MOP. Under this new formulation, a set of equally acceptable Pareto optimal solutions is provided, rather than a single solution that optimizes a specific objective. We can distinguish between two tasks, namely, i) finding a set of Pareto optimal solutions, and ii) choosing the most preferred solution out of this set in a decision-making procedure. Second, we provide two types of trade-off parameters, namely, scaling and weighting parameters. The scaling parameters define a family of MOPs (i.e., Pareto-2(SD,ONR) and Pareto-3(SD,IR,NR) MOPs). These parameters determine the respective MOP to be optimized. The weighting parameters are used to select a preferred solution out of a set of Pareto solutions in a decision-making procedure. Third, we provide a list of considerations for the designer to control the binaural cue preservation, the speech distortion (SD), and the noise reduction (NR). Fourth, we show that a wide range of well-known binaural beamformers are all sub-sets of the proposed framework, namely, they all lie on the respective Pareto frontier. Fifth, we establish a link between two well-known beamforming families, namely, the binaural MVDR-based and the MWF-based beamformers, as an example of the usage of the Pareto formulation.

The paper structure follows. In Section II, we define the considered signal model, the interaural criteria, the MSE cost functions, and the performance measures. In Section III, the Pareto MOP formulation is introduced, a multi-criteria is proposed for the binaural problem, and the binaural beamformer satisfying the MOP, referred to as the Pareto-BMWF beamformer, is derived. The binaural cue preservation properties of the proposed Pareto-BMWF are analyzed in Section IV. In Section V, several known binaural MWF-based and MVDR-based beamformers are shown to be special cases of the proposed beamformer, and insights into the relation between the binaural MWF and the binaural MVDR beamformer are given. In Section VI, experiments with real signals that demonstrate the performance of the proposed Pareto-BMWF beamformer when estimation errors are present are described. In Section VII, a discussion is provided and we conclude the paper.

II. PROBLEM FORMULATION

In this section, we introduce the considered signal model (Section II-A), define the binaural cues (Section II-B), present the MSE cost functions (Section II-C) and the performance measures (Section II-D).

A. Microphone and Output Signals

1) General Case: Consider an acoustic scenario consisting of target and interfering sources in a noisy and reverberant environment. The sources are received by two fully connected hearing aid devices consisting of a microphone array with M_L microphones on the left hearing aid and M_R microphones on the right hearing aid, where $M = M_L + M_R$ denotes the total number of microphones. The received signal in the short-time Fourier transform (STFT) domain can be formulated as an M-dimensional vector $\boldsymbol{y}(t,k) = [y_{L,1}(t,k), \ldots, y_{L,M_L}(t,k), \ y_{R,1}(t,k), \ldots, y_{R,M_R}(t,k)]^T$, which can be written as

$$\boldsymbol{y}(t,k) = \boldsymbol{x}(t,k) + \boldsymbol{u}(t,k) + \boldsymbol{n}(t,k)$$
$$= \boldsymbol{x}(t,k) + \boldsymbol{v}(t,k), \qquad (1)$$

where k denotes the frequency index and t the frame index, and $\boldsymbol{x}(t,k)$, $\boldsymbol{u}(t,k)$, and $\boldsymbol{n}(t,k)$ denote the received target source component, the received directional interfering (undesired) source component, and the received background noise component, respectively. $\boldsymbol{v}(t,k) = \boldsymbol{u}(t,k) + \boldsymbol{n}(t,k)$ is defined as the overall noise component as received by the microphones, i.e., the directional interfering source component plus the background noise component. The spatial correlation matrices of the target source, interfering source, and background noise components' $\boldsymbol{R}_X$, $\boldsymbol{R}_U$ and $\boldsymbol{R}_N$, are defined as

$$\boldsymbol{R}_X(t,k) = \mathcal{E}\{\boldsymbol{x}(t,k)\boldsymbol{x}^H(t,k)\},$$
$$\boldsymbol{R}_U(t,k) = \mathcal{E}\{\boldsymbol{u}(t,k)\boldsymbol{u}^H(t,k)\},$$
$$\boldsymbol{R}_N(t,k) = \mathcal{E}\{\boldsymbol{n}(t,k)\boldsymbol{n}^H(t,k)\}, \qquad (2)$$

where $\mathcal{E}\{\cdot\}$ denotes the expectation operator. Assuming statistical independence between the components in (1), the spatial correlation matrix of the microphone signals $\boldsymbol{R}_Y$ can be written as

$$\boldsymbol{R}_Y(t,k) = \mathcal{E}\{\boldsymbol{y}(t,k)\boldsymbol{y}^H(t,k)\}$$
$$= \boldsymbol{R}_X(t,k) + \underbrace{\boldsymbol{R}_U(t,k) + \boldsymbol{R}_N(t,k)}_{\boldsymbol{R}_V(t,k)}, \qquad (3)$$

with $\boldsymbol{R}_V$ the spatial correlation matrix of the overall noise component.

Let m_L and m_R be the indices of the left and right reference microphones, respectively (usually selected as the microphones closest to the ears in order to best reflect the binaural cues). The respective reference microphone signals at the left and the right hearing aids are given by

$$y_L(t,k) = \boldsymbol{e}_L^H \boldsymbol{y}(t,k), \quad y_R(t,k) = \boldsymbol{e}_R^H \boldsymbol{y}(t,k), \qquad (4)$$

where $\boldsymbol{e}_L$ and $\boldsymbol{e}_R$ are M-dimensional vectors with '1' in the m_Lth and m_Rth component, respectively, and '0' elsewhere. The output signals on the left and the right hearing aids, z_L and z_R respectively, are obtained by applying the left and the right

beamformers to all microphone signals from both hearing aids, i.e.,

$$z_L(t,k) = \boldsymbol{w}_L^H(t,k)\boldsymbol{y}(t,k), \quad z_R(t,k) = \boldsymbol{w}_R^H(t,k)\boldsymbol{y}(t,k), \quad (5)$$

where $\boldsymbol{w}_L(t,k)$ and $\boldsymbol{w}_R(t,k)$ are M-dimensional complex-valued weight vectors for the left and the right hearing aid, respectively. Furthermore, we define the $2M$-dimensional complex-valued stacked weight vector $\boldsymbol{w}(t,k)$ as

$$\boldsymbol{w}(t,k) = \begin{bmatrix} \boldsymbol{w}_L(t,k) \\ \boldsymbol{w}_R(t,k) \end{bmatrix}. \quad (6)$$

Henceforth, t and k are omitted for the sake of brevity.

2) Special Case: Dual Source Scenario: In this section, we consider a common scenario consisting of one target source, one interfering source (e.g., competing speakers), and background noise, which can be directional, non-directional, or a combination thereof (although commonly a diffuse noise).

We assume a deterministic characterization of acoustic impulse responses (AIRs). The M-dimensional vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ denote the acoustic transfer function (ATF) vectors from the (point) sources to the microphones. We note that the AIRs vary in time due to the movements of the sources, the microphones, or other objects in the environment. Nevertheless, we assume such movements are small, and hence, the ATFs $\boldsymbol{a}$ and $\boldsymbol{b}$ can be approximated as deterministic time-invariant.

Under the above assumptions, the target and the interfering source components can be modeled as

$$\boldsymbol{x} = s_X \boldsymbol{a}, \quad \boldsymbol{u} = s_U \boldsymbol{b}, \quad (7)$$

where s_X and s_U denote the target and interfering source signals, respectively. In this case, the correlation matrices $\boldsymbol{R}_X$ and $\boldsymbol{R}_U$ are rank-1 matrices, i.e.,

$$\boldsymbol{R}_X = P_X \boldsymbol{a}\boldsymbol{a}^H, \quad \boldsymbol{R}_U = P_U \boldsymbol{b}\boldsymbol{b}^H. \quad (8)$$

with $P_X = \mathcal{E}\{|s_X|^2\}$ and $P_U = \mathcal{E}\{|s_U|^2\}$ denoting the power spectral density (PSD) of the target and interfering source components, respectively.

The relative transfer functions (RTFs) of the target source and the interfering source between the reference microphones on the left and the right hearing aids are defined as the ratio of the ATFs, i.e.,

$$\text{RTF}_{X,\text{IN}} = \frac{a_L}{a_R}, \quad \text{RTF}_{U,\text{IN}} = \frac{b_L}{b_R}. \quad (9)$$

B. Binaural Cues

1) General Case: The input and output interaural transfer function (ITF) of the target and interfering source components are defined as the ratio of the components at the left and right hearing aids [11], i.e.,

$$\text{ITF}_{X,\text{IN}} = \frac{x_L}{x_R} = \frac{\boldsymbol{e}_L^H \boldsymbol{x}}{\boldsymbol{e}_R^H \boldsymbol{x}}, \quad \text{ITF}_{X,\text{OUT}} = \frac{\boldsymbol{w}_L^H \boldsymbol{x}}{\boldsymbol{w}_R^H \boldsymbol{x}},$$

$$\text{ITF}_{U,\text{IN}} = \frac{u_L}{u_R} = \frac{\boldsymbol{e}_L^H \boldsymbol{u}}{\boldsymbol{e}_R^H \boldsymbol{u}}, \quad \text{ITF}_{U,\text{OUT}} = \frac{\boldsymbol{w}_L^H \boldsymbol{u}}{\boldsymbol{w}_R^H \boldsymbol{u}}. \quad (10)$$

The ITF is a complex-valued frequency-dependent scalar, from which the ILD and the ITD binaural cues can be computed

as [11][1]

$$\text{ILD} = 20 \log_{10}(|\text{ITF}|), \quad \text{ITD} = \frac{\angle(\text{ITF})}{\omega}, \quad (11)$$

with $\angle$ denoting the unwrapped phase and ω the radian frequency. The interaural phase difference (IPD) is defined as the phase of the ITF, i.e.,

$$\text{IPD} = \angle(\text{ITF}). \quad (12)$$

In the following, because of the relation in (11), we consider that the ITF preservation capabilities of the examined filters are equivalent to their binaural cue preservation capabilities for the sake of brevity.

2) Special Case: Dual Source Scenario: For a specific single source, the input ITF of either the target or the interfering source is independent of the actual input signal and equals the RTF between the reference microphones at the left and the right hearing aids as defined in (9), i.e.,

$$\text{ITF}_{X,\text{IN}} = \frac{a_L}{a_R}, \quad \text{ITF}_{U,\text{IN}} = \frac{b_L}{b_R}. \quad (13)$$

Similarly, the output ITFs of the target and the interfering sources are equal to the output RTFs of the target and the interfering sources, which are defined as the ratio of the filtered components at the left and the right hearing aids, i.e.,

$$\text{ITF}_{X,\text{OUT}} = \frac{\boldsymbol{w}_L^H \boldsymbol{a}}{\boldsymbol{w}_R^H \boldsymbol{a}}, \quad \text{ITF}_{U,\text{OUT}} = \frac{\boldsymbol{w}_L^H \boldsymbol{b}}{\boldsymbol{w}_R^H \boldsymbol{b}}. \quad (14)$$

For practical implementations, for a single source, the input and output ITFs of the target and interfering sources can be estimated from the spatial correlation matrix [11], i.e.,[2]

$$\text{ITF}_{X,\text{IN}} = \frac{\boldsymbol{e}_L^H \boldsymbol{R}_X \boldsymbol{e}_L}{\boldsymbol{e}_R^H \boldsymbol{R}_X \boldsymbol{e}_L}, \quad \text{ITF}_{X,\text{OUT}} = \frac{\boldsymbol{w}_L^H \boldsymbol{R}_X \boldsymbol{w}_L}{\boldsymbol{w}_R^H \boldsymbol{R}_X \boldsymbol{w}_L},$$

$$\text{ITF}_{U,\text{IN}} = \frac{\boldsymbol{e}_L^H \boldsymbol{R}_U \boldsymbol{e}_L}{\boldsymbol{e}_R^H \boldsymbol{R}_U \boldsymbol{e}_L}, \quad \text{ITF}_{U,\text{OUT}} = \frac{\boldsymbol{w}_L^H \boldsymbol{R}_U \boldsymbol{w}_L}{\boldsymbol{w}_R^H \boldsymbol{R}_U \boldsymbol{w}_L}. \quad (15)$$

For example, using (15), and since the spatial correlation matrices are rank-1, the input ITFs of the target and interfering sources are equal to the respective input RTFs, i.e.,

$$\text{ITF}_{X,\text{IN}} = \frac{\boldsymbol{e}_L^H P_X \boldsymbol{a}\boldsymbol{a}^H \boldsymbol{e}_L}{\boldsymbol{e}_R^H P_X \boldsymbol{a}\boldsymbol{a}^H \boldsymbol{e}_L} = \frac{a_L a_L^*}{a_R a_L^*} = \frac{a_L}{a_R},$$

$$\text{ITF}_{U,\text{IN}} = \frac{\boldsymbol{e}_L^H P_U \boldsymbol{b}\boldsymbol{b}^H \boldsymbol{e}_L}{\boldsymbol{e}_R^H P_U \boldsymbol{b}\boldsymbol{b}^H \boldsymbol{e}_L} = \frac{b_L b_L^*}{b_R b_L^*} = \frac{b_L}{b_R}. \quad (16)$$

Hence, the ITF is equivalent to the RTF. Henceforth, the ITF is referred to as the RTF for the sake of clarity.

C. Mean Square Error Objectives

In this study, we aim at minimizing three criteria: the target SD, (partial) IR, and (partial) NR terms. In this section, useful definitions related to these criteria are provided.

---

[1]Please note that in some publications the ITD is defined as derived from the phase of the ITF [31], i.e., $\text{ITD} = -\frac{d}{d\omega} \angle(\text{ITF})$. We use the ITD as defined in [11].

[2]Note that (15) is also used in this study in the general case where the rank of correlation matrices is larger than one.

We define the signal-based cost function $J_x(\boldsymbol{w})$ for the target SD term as the MSE between the target source component in the reference microphone signals and the output signals, i.e.,

$$J_x(\boldsymbol{w}) = \mathcal{E}\left\{\left\|\begin{bmatrix} x_L - \boldsymbol{w}_L^H \boldsymbol{x} \\ x_R - \boldsymbol{w}_R^H \boldsymbol{x} \end{bmatrix}\right\|^2\right\}. \tag{17}$$

In addition, we define an equivalent transfer function (TF)-based cost function $T_x(\boldsymbol{w})$ for the SD term using the ATF vector of the target source, i.e.,

$$T_x(\boldsymbol{w}) = \left\{\left\|\begin{bmatrix} a_L - \boldsymbol{w}_L^H \boldsymbol{a} \\ a_R - \boldsymbol{w}_R^H \boldsymbol{a} \end{bmatrix}\right\|^2\right\}. \tag{18}$$

Note that, while MWF-based beamformers require signal-based cost functions for the SD term, distortionless beamformers require a TF-based cost function of the target source for the SD term (Section V).

For a single source case, the target source is modeled in (7) as $\boldsymbol{x} = s_X \boldsymbol{a}$, and hence, the correlation matrix $\boldsymbol{R}_X$ in (8) is a rank-1 matrix. Substituting (7) in (17), the signal-based cost function $J_x(\boldsymbol{w})$ can be written as

$$J_x(\boldsymbol{w}) = \mathcal{E}\left\{\left\|\begin{bmatrix} s_X a_L - \boldsymbol{w}_L^H s_X \boldsymbol{a} \\ s_X a_R - \boldsymbol{w}_R^H s_X \boldsymbol{a} \end{bmatrix}\right\|^2\right\}$$

$$= \mathcal{E}\left\{\|s_X\|^2\right\} \left\{\left\|\begin{bmatrix} a_L - \boldsymbol{w}_L^H \boldsymbol{a} \\ a_R - \boldsymbol{w}_R^H \boldsymbol{a} \end{bmatrix}\right\|^2\right\}. \tag{19}$$

Since $P_X = \mathcal{E}\{|s_X|^2\}$, the signal-based cost function $J_x(\boldsymbol{w})$ is equivalent to the TF-based cost function $T_x(\boldsymbol{w})$, except for a scaling factor, which is equal to the PSD of the signal source, i.e.,

$$J_x(\boldsymbol{w}) = P_X T_x(\boldsymbol{w}). \tag{20}$$

We define the signal-based cost function $J_{u,\eta_u}(\boldsymbol{w})$ for the IR term as

$$J_{u,\eta_u}(\boldsymbol{w}) = \mathcal{E}\left\{\left\|\begin{bmatrix} \eta_u u_L - \boldsymbol{w}_L^H \boldsymbol{u} \\ \eta_u u_R - \boldsymbol{w}_R^H \boldsymbol{u} \end{bmatrix}\right\|^2\right\}, \tag{21}$$

where $0 \leq \eta_u \leq 1$ denotes the interference *scaling parameter*, which controls the amount of IR. Further, we define an equivalent TF-based cost function $T_{u,\eta_u}(\boldsymbol{w})$ for the IR term using the ATF vector of the interfering source, i.e.,

$$T_{u,\eta_u}(\boldsymbol{w}) = \left\{\left\|\begin{bmatrix} \eta_u b_L - \boldsymbol{w}_L^H \boldsymbol{b} \\ \eta_u b_R - \boldsymbol{w}_R^H \boldsymbol{b} \end{bmatrix}\right\|^2\right\}. \tag{22}$$

For the single source case, the relation between the signal-based cost function $J_{u,\eta_u}(\boldsymbol{w})$ and the TF-based cost function $T_{u,\eta_u}(\boldsymbol{w})$ for the interfering source is, similar to the above relation for the target source, given by,

$$J_{u,\eta_u}(\boldsymbol{w}) = P_U T_{u,\eta_u}(\boldsymbol{w}), \tag{23}$$

such that the signal-based cost function $J_{u,\eta_u}(\boldsymbol{w})$ is equivalent to the TF-based cost function $T_{u,\eta_u}(\boldsymbol{w})$, up to a scaling factor, which is equal to the PSD of the interfering source.

We define the signal-based cost function $J_{n,\eta_n}(\boldsymbol{w})$ for the NR term as

$$J_{n,\eta_n}(\boldsymbol{w}) = \mathcal{E}\left\{\left\|\begin{bmatrix} \eta_n n_L - \boldsymbol{w}_L^H \boldsymbol{n} \\ \eta_n n_R - \boldsymbol{w}_R^H \boldsymbol{n} \end{bmatrix}\right\|^2\right\}, \tag{24}$$

where $0 \leq \eta_n \leq 1$ denotes the background noise *scaling parameter*, which controls the amount of noise reduction.

Furthermore, we define the cost function $J_{v,\eta_v}(\boldsymbol{w})$ for the overall noise reduction (ONR) as

$$J_{v,\eta_v}(\boldsymbol{w}) = \mathcal{E}\left\{\left\|\begin{bmatrix} \eta_v v_L - \boldsymbol{w}_L^H \boldsymbol{v} \\ \eta_v v_R - \boldsymbol{w}_R^H \boldsymbol{v} \end{bmatrix}\right\|^2\right\}, \tag{25}$$

where $0 \leq \eta_v \leq 1$ denotes the overall noise *scaling parameter*, which controls the amount of overall noise reduction (i.e., interfering source plus background noise).

Remark 1: When setting the *scaling parameters* η_u, η_n, and η_v, different important aspects should be considered, e.g., those based on the desired SIR and SNR improvement and the effect of RTF estimation errors. This may improve situational awareness and, consequently, enhance the preservation of the spatial cues. This set of parameters can be determined by the user based on personal perceptual preferences rather than tuned by an optimization process maximizing a particular measure of performance.

Remark 2: All the proposed MSE cost functions are, by definition, non-negative values.

Remark 3: When the cost function for the target source $J_x(\boldsymbol{w})$ (or equivalently, $T_x(\boldsymbol{w})$) is equal to zero, the filters must satisfy a constraint set, namely, $\boldsymbol{w} \in H_x(\boldsymbol{w})$, with

$$H_x(\boldsymbol{w}) = \left\{\boldsymbol{w} \in \mathbb{C}^M : \boldsymbol{w}_L^H \boldsymbol{a} = a_L, \boldsymbol{w}_R^H \boldsymbol{a} = a_R\right\}. \tag{26}$$

This leads to a family of distortionless beamformers, as elaborated in Section V-B.

Similarly, when the cost function for the interfering source $J_{u,\eta_u}(\boldsymbol{w})$ (or equivalently, $T_{u,\eta_u}(\boldsymbol{w})$) is equal to zero, the filters must satisfy a constraint set, $\boldsymbol{w} \in H_{u,\eta_u}(\boldsymbol{w})$ with

$$H_{u,\eta_u}(\boldsymbol{w}) = \left\{\boldsymbol{w} \in \mathbb{C}^M : \boldsymbol{w}_L^H \boldsymbol{b} = \eta_u b_L, \boldsymbol{w}_R^H \boldsymbol{b} = \eta_u b_R\right\}. \tag{27}$$

This leads to a family of null-steering beamformers, as elaborated in Section V-A-3 and Section V-B.

D. Performance Measures

In this section, useful performance measures, which are used in Section VI, are defined.

The *narrow-band binaural SD* is defined as the average of the left and right target SD terms (cf. Eq. (17)) normalized by the average input PSD of the left and right target component on the reference microphones, i.e.,

$$\text{SD} = \frac{\mathcal{E}\left\{\|x_L - \boldsymbol{w}_L^H \boldsymbol{x}\|^2\right\} + \mathcal{E}\left\{\|x_R - \boldsymbol{w}_R^H \boldsymbol{x}\|^2\right\}}{\mathcal{E}\left\{\|x_L\|^2\right\} + \mathcal{E}\left\{\|x_R\|^2\right\}}. \tag{28}$$

The binaural SD can be rewritten as

$$\text{SD} = \frac{\mathcal{E}\left\{\|x_L - \boldsymbol{w}_L^H \boldsymbol{x}\|^2\right\} + \mathcal{E}\left\{\|x_R - \boldsymbol{w}_R^H \boldsymbol{x}\|^2\right\}}{\boldsymbol{e}_L^H \boldsymbol{R}_X \boldsymbol{e}_L + \boldsymbol{e}_R^H \boldsymbol{R}_X \boldsymbol{e}_R}. \tag{29}$$

The *narrow-band binaural input signal-to-interference-and-noise ratio (SINR)* is defined as the ratio of the average input PSDs of the target source and the overall noise components (interfering source plus background noise) in the reference microphones, i.e.,

$$\text{SINR}^{\text{in}} = \frac{e_L^H R_X e_L + e_R^H R_X e_R}{e_L^H R_V e_L + e_R^H R_V e_R}. \tag{30}$$

The *narrow-band binaural output SINR* is defined as the ratio of the average output PSDs of the target source and the overall noise components (interfering source plus background noise) in the left and the right hearing aid, i.e., [14]

$$\text{SINR}^{\text{out}} = \frac{w_L^H R_X w_L + w_R^H R_X w_R}{w_L^H R_V w_L + w_R^H R_V w_R}. \tag{31}$$

III. THE MULTI OPTIMIZATION PROBLEM

In this section, the multi-objective problem (MOP) is introduced. First, the mathematical foundations of the MOP are laid (Section III-A). Second, the considered binaural problem is expressed in a MOP formulation, and a novel binaural beamformer denoted Pareto-BMWF is derived, constituting a Pareto optimal set of filters solving the MOP (Section III-B).

A. Mathematical Formulation

MOP is a field of multiple criteria decision making that involves the simultaneous optimization of more than one objective function. This section provides a brief description of the MOP that is used in the next sections, following the description in [32]–[34]. The interested reader is referred to [30], [35], [36].

A single-objective problem (SOP) for the binaural system considered in this study is given by finding the binaural filter solution w that minimizes a single cost function $J(w)$, i.e.,

$$\min_{w} J(w). \tag{32}$$

The optimization technique leads to a unique optimal filter.

The general MOP can be described as a simultaneous optimization of multiple cost functions, i.e.,

$$\min_{w} \text{Pareto } \mathcal{C}(w) \tag{33}$$

with

$$\mathcal{C}(w) = [J_1(w), J_2(w), \ldots, J_I(w)], \tag{34}$$

where I is the number of cost functions. The binaural filter w is a vector of design variables. $\mathcal{C}(w)$ is a vector-valued global objective function, or the generalized cost function, such that it represents a set of criteria. Each $J_i(w)$ denotes an objective associated with a different cost function. w_i^* is a filter solution that minimizes the cost function $J_i(w)$. The feasible design space $\mathcal{W}$ (frequently called the feasible decision space) is defined as the set of all possible solutions w. The feasible criterion space $\mathcal{Z}$ (also called the feasible cost space or the attainable set) is defined as the set $\mathcal{Z} = \{\mathcal{C}(w)|w \in \mathcal{W}\}$. Each point in the design space maps to a point (vector-valued) in the criterion space, but the reverse may not be true. We attribute *attainability* to a point

(vector-valued) in the criterion space that maps to a point in the design space.

If there is a feasible solution that minimizes all criteria, then a single solution is obtained, similar to the SOP. However, it is typically impossible to find a unique solution that minimizes all criteria. In these situations, the notion of Pareto optimality becomes of paramount importance. A set of Pareto-optimal solutions comprises all solutions that minimize each cost function individually and also the solutions that trade off these cost functions. The Pareto optimality is defined as follows.

Definition 1: A filter solution w^* dominates another filter solution w if i) w^* is no worse than w in all objectives and ii) w^* is strictly better than w in at least one objective.

Definition 2: A filter solution $w^* \in \mathcal{W}$ is Pareto optimal solution iff there does not exist another filter solution $w \in \mathcal{W}$, such that $J_i(w) \leq J_i(w^*)$ for all $i = 1, 2, \ldots, I$ and $J_j(w) < J_j(w^*)$ for at least one index j (cost function). In other words, a filter solution w^* is Pareto optimal if it is not dominated by any other filter solution.

Definition 3: All the Pareto optimal filter solutions solve the MOP and lie on the boundary of the feasible criterion space $\mathcal{Z}$ [37]. The set of Pareto solutions constitutes the Pareto *frontier*.

This means that a filter is a Pareto solution if no other filter exists that improves at least one cost function without leading to a degradation in another cost function. We note that for any given MOP, there may be an infinite number of Pareto optimal solutions constituting the Pareto optimal set.

The Pareto frontier is the optimal trade-off between multiple objectives. As in our study, all cost functions are convex by definition (see Section II-C Remark 2), the local Pareto solution is also global Pareto optimal. Then, one does not need to discuss particular cases because all the solutions on the frontier are optimal. As the "operator" of the optimization process that determines a specific solution, it is guaranteed that any selected point on the frontier gives an optimal trade-off between the cost functions. In that sense, every Pareto optimal point is an equally acceptable solution of the MOP.

We use the so-called weighted-sum or *scalarization* method, as described in [33], to compute the Pareto set. The MOP is solved by combining its multiple objectives into one single-objective generalized cost function. This is defined as

$$J(w) = \lambda_1 J_1(w) + \lambda_2 J_2(w) + \ldots + \lambda_I J_I(w)$$

$$\text{s.t.} \sum_{i=1}^{I} \lambda_i = 1, \tag{35}$$

where $\lambda_i, \ i = 1, 2, \ldots, I$ are defined as the *weighting parameters* that provide a trade-off between the cost function terms.

We note that the minimization of the single-objective generalized cost function $J(w)$ is sufficient for finding a Pareto optimal solution if $J(w)$ increases monotonically with respect to each cost function [32]. That is, any filter that solves the generalized cost function belongs to the Pareto frontier.

It is generally desirable to obtain one point as a solution out of the Pareto optimal set. This can be accomplished using a *decision-making* procedure [34], based on considerations that

are independent of the MOP (this can be accomplished by, e.g., by setting the weighting parameters). Hence, a second stage in the MOP framework is to define a decision-making procedure that selects a preferred solution from the Pareto frontier based on a list of considerations (e.g., required SD, SIR, NR, or binaural cue preservation requirements).

B. Binaural Multiple Objective Problem

In this section, we apply the MOP formulation to the binaural problems at hand.

1) Pareto Three-Objectives Problem: In Section II-C, we defined the MSE cost functions for SD, IR, and NR. The objective of the examined problem is to minimize simultaneously the three cost functions. The MOP can be described in mathematical terms as

$$\min_{\boldsymbol{w}} \text{Pareto } \mathcal{C}_{\text{Pareto-3}}(\boldsymbol{w}) \qquad (36)$$

with

$$\mathcal{C}_{\text{Pareto-3}}(\boldsymbol{w}) = [J_x(\boldsymbol{w}), J_{u,\eta_u}(\boldsymbol{w}), J_{n,\eta_n}(\boldsymbol{w})]. \qquad (37)$$

Since the problem optimizes three cost functions, it is referred to as a Pareto-3 problem.

To calculate the Pareto frontier, we use the scalarization method. The generalized MSE cost function consists of a weighted sum of the three cost functions (17), (21), and (24), i.e.,

$$J_{\text{Pareto-3}(\eta_u,\eta_n)}(\boldsymbol{w}) = \lambda_x J_x(\boldsymbol{w}) + \lambda_u J_{u,\eta_u}(\boldsymbol{w}) + \lambda_n J_{n,\eta_n}(\boldsymbol{w})$$
$$\text{s.t. } \lambda_x + \lambda_u + \lambda_n = 1, \qquad (38)$$

where the target, interference, and background noise *weighting parameters* $0 \leq \lambda_x \leq 1$, $0 \leq \lambda_u \leq 1$, and $0 \leq \lambda_n \leq 1$ provide a trade-off between SD, IR, and NR MSE terms. The optimal set of solutions can be found by seeking minima through varying the weighting parameters λ_x and λ_u, where $\lambda_n = 1 - \lambda_x - \lambda_u$. This is possible, since the solution of the generalized cost function for any λ_x and λ_u corresponds to a particular filter, which forms the Pareto frontier. In the second stage, the decision-making procedure will select the preferred optimal solution from the Pareto frontier, by determining the weighting parameters.

2) Pareto Two-Objectives Problem: Let us examine a setting where $\lambda_v = \lambda_u = \lambda_n$ and $\eta_v = \eta_u = \eta_n$ such that the interfering source and the background noise are similarly treated. The generalized MSE cost function in (38) is then equal to

$$J_{\text{Pareto-3}(\eta_u,\eta_n)}(\boldsymbol{w}) = \lambda_x J_x(\boldsymbol{w}) + \lambda_v (J_{u,\eta_u}(\boldsymbol{w}) + J_{n,\eta_n}(\boldsymbol{w}))$$
$$\text{s.t. } \lambda_x + \lambda_v = 1. \qquad (39)$$

Assuming statistical independence between $\boldsymbol{u}$ and $\boldsymbol{n}$, it can be shown that the cost function $J_{v,\eta_v}(\boldsymbol{w})$ for the ONR is equal to the sum of the signal-based cost function $J_{u,\eta_u}(\boldsymbol{w})$ for the IR term and the signal-based cost function $J_{n,\eta_n}(\boldsymbol{w})$ for the NR term, i.e.,

$$J_{v,\eta_v}(\boldsymbol{w}) = J_{u,\eta_u}(\boldsymbol{w}) + J_{n,\eta_n}(\boldsymbol{w}). \qquad (40)$$

Hence, the objective of the examined problem is to simultaneously minimize two cost functions (referred to as a Pareto-2 problem):

$$\min_{\boldsymbol{w}} \text{Pareto } \mathcal{C}_{\text{Pareto-2}}(\boldsymbol{w}) \qquad (41)$$

with

$$\mathcal{C}_{\text{Pareto-2}}(\boldsymbol{w}) = [J_x(\boldsymbol{w}), J_{v,\eta_v}(\boldsymbol{w})]. \qquad (42)$$

The generalized cost function is equal to

$$J_{\text{Pareto-2}(\eta_v)}(\boldsymbol{w}) = \lambda_x J_x(\boldsymbol{w}) + (1 - \lambda_x)J_{v,\eta_v}(\boldsymbol{w}), \qquad (43)$$

where $\lambda_v = 1 - \lambda_x$.

3) Pareto Three-Objectives Filter Decomposition: The filters minimizing the cost function in (38) can be computed as

$$\boldsymbol{w}_L = \boldsymbol{R}_\lambda^{-1} \boldsymbol{R}_\eta \boldsymbol{e}_L, \ \boldsymbol{w}_R = \boldsymbol{R}_\lambda^{-1} \boldsymbol{R}_\eta \boldsymbol{e}_R, \qquad (44)$$

with

$$\boldsymbol{R}_\lambda = \lambda_x \boldsymbol{R}_X + \lambda_u \boldsymbol{R}_U + \lambda_n \boldsymbol{R}_N \qquad (45)$$

and

$$\boldsymbol{R}_\eta = \lambda_x \boldsymbol{R}_X + \lambda_u \eta_u \boldsymbol{R}_U + \lambda_n \eta_n \boldsymbol{R}_N. \qquad (46)$$

The filters are referred to as the Pareto-BMWF, with $\lambda_n = 1 - \lambda_x - \lambda_u$. Further, the left and right filters for the Pareto-BMWF can be written in a unified stacked vector as

$$\boldsymbol{w} = \boldsymbol{R}_G^{-1} \boldsymbol{r}_G, \qquad (47)$$

with

$$\boldsymbol{R}_G = \begin{bmatrix} \boldsymbol{R}_\lambda & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R}_\lambda \end{bmatrix}, \qquad (48)$$

and

$$\boldsymbol{r}_G = \begin{bmatrix} \boldsymbol{R}_\eta \boldsymbol{e}_L \\ \boldsymbol{R}_\eta \boldsymbol{e}_R \end{bmatrix}. \qquad (49)$$

By substituting the rank-1 correlation matrices $\boldsymbol{R}_X$ and $\boldsymbol{R}_U$ from (8) into (44), we obtain filters that can be decomposed as

$$\boldsymbol{w}_L = \boldsymbol{w}_{X,L} + \eta_u \boldsymbol{w}_{U,L} + \eta_n \boldsymbol{w}_{N,L},$$
$$\boldsymbol{w}_R = \boldsymbol{w}_{X,R} + \eta_u \boldsymbol{w}_{U,R} + \eta_n \boldsymbol{w}_{N,R}, \qquad (50)$$

with

$$\boldsymbol{w}_{X,L} = \lambda_x \boldsymbol{R}_\lambda^{-1} P_X \boldsymbol{a} a_L^* \quad \boldsymbol{w}_{X,R} = \lambda_x \boldsymbol{R}_\lambda^{-1} P_X \boldsymbol{a} a_R^*$$
$$\boldsymbol{w}_{U,L} = \lambda_u \boldsymbol{R}_\lambda^{-1} P_U \boldsymbol{b} b_L^* \quad \boldsymbol{w}_{U,R} = \lambda_u \boldsymbol{R}_\lambda^{-1} P_U \boldsymbol{b} b_R^*$$
$$\boldsymbol{w}_{N,L} = \lambda_n \boldsymbol{R}_\lambda^{-1} \boldsymbol{R}_N \boldsymbol{e}_L \quad \boldsymbol{w}_{N,R} = \lambda_n \boldsymbol{R}_\lambda^{-1} \boldsymbol{R}_N \boldsymbol{e}_R. \qquad (51)$$

The obtained filters are decomposed into a sum of three filters, related to the target, interference, and noise sources. This decomposition can be interpreted as a source separation procedure that first calculates three outputs and then, remixes the outputs by the trade-off parameters.

Both (λ_x, λ_u) and (η_u, η_n) control the Pareto-3(η_u,η_n) MOP. The scaling parameters (η_u, η_n) control the amount of interference and noise reduction required (internally) in the IR and NR MSE terms, respectively. Various Pareto-3(η_u,η_n) frontiers are obtained as a function of these scaling parameters. On the other hand, the target and interference weighting parameters (λ_x, λ_u) provide a trade-off between the SD, IR, and NR MSE terms,

such that they give us a tool for selecting a point on the specific Pareto-3(η_u, η_n) frontier in the decision-making procedure.

From (50), it is evident that the Pareto-3 filters are a sum of the target filters $\boldsymbol{w}_{X,L}$ and $\boldsymbol{w}_{X,R}$, the interference filters $\boldsymbol{w}_{U,L}$ and $\boldsymbol{w}_{U,R}$ (weighted with η_u), and the noise filters $\boldsymbol{w}_{N,L}$ and $\boldsymbol{w}_{N,R}$ (weighted with η_n). For $\eta_u = 1$, and $\eta_n = 1$, using (44), the filters are reduced to the identity filters (i.e., $\boldsymbol{w}_L = \boldsymbol{e}_L$ and $\boldsymbol{w}_R = \boldsymbol{e}_R$) since $\boldsymbol{R}_\eta = \boldsymbol{R}_\lambda$, such that the output of each filter is equal to the respective noisy reference microphone signal.

4) Pareto Two-Objectives Filter Decomposition: The filters minimizing the cost function in (43) are given by

$$\boldsymbol{w}_L = \bar{\boldsymbol{R}}_\lambda^{-1} \bar{\boldsymbol{R}}_\eta \boldsymbol{e}_L, \; \boldsymbol{w}_R = \bar{\boldsymbol{R}}_\lambda^{-1} \bar{\boldsymbol{R}}_\eta \boldsymbol{e}_R, \tag{52}$$

with

$$\bar{\boldsymbol{R}}_\lambda = \lambda_x \boldsymbol{R}_X + \lambda_v \boldsymbol{R}_V; \quad \bar{\boldsymbol{R}}_\eta = \lambda_x \boldsymbol{R}_X + \lambda_v \eta_v \boldsymbol{R}_V, \tag{53}$$

and a stacked vector can be written by substituting $\boldsymbol{R}_\lambda$ and $\boldsymbol{R}_\eta$ with $\bar{\boldsymbol{R}}_\lambda$ and $\bar{\boldsymbol{R}}_\eta$ in (47), i.e.,

$$\boldsymbol{w} = \bar{\boldsymbol{R}}_G^{-1} \bar{\boldsymbol{r}}_G, \tag{54}$$

with

$$\bar{\boldsymbol{R}}_G = \begin{bmatrix} \bar{\boldsymbol{R}}_\lambda & \boldsymbol{0} \\ \boldsymbol{0} & \bar{\boldsymbol{R}}_\lambda \end{bmatrix}, \tag{55}$$

and

$$\bar{\boldsymbol{r}}_G = \begin{bmatrix} \bar{\boldsymbol{R}}_\eta \boldsymbol{e}_L \\ \bar{\boldsymbol{R}}_\eta \boldsymbol{e}_R \end{bmatrix}. \tag{56}$$

By substituting the rank-1 correlation matrix $\boldsymbol{R}_X$ from (8) into (52), we obtain filters

$$\boldsymbol{w}_L = (1 - \eta_v) \bar{\boldsymbol{w}}_{X,L} + \eta_v \boldsymbol{e}_L,$$
$$\boldsymbol{w}_R = (1 - \eta_v) \bar{\boldsymbol{w}}_{X,R} + \eta_v \boldsymbol{e}_R, \tag{57}$$

where $\bar{\boldsymbol{w}}_{X,L}$ and $\bar{\boldsymbol{w}}_{X,R}$ are the Pareto-2($\eta_v = 0$) filters, i.e.,

$$\bar{\boldsymbol{w}}_{X,L} = \lambda_x \bar{\boldsymbol{R}}_\lambda^{-1} P_X \boldsymbol{a} \boldsymbol{a}_L^*,$$
$$\bar{\boldsymbol{w}}_{X,R} = \lambda_x \bar{\boldsymbol{R}}_\lambda^{-1} P_X \boldsymbol{a} \boldsymbol{a}_R^*. \tag{58}$$

The solution for this problem defines a Pareto frontier for the Pareto-2(η_v) MOP.

Both λ_x and η_v control the Pareto-2(η_v) MOP. The target weighting parameter λ_x provides a trade-off between SD and ONR MSE terms. For $\eta_v = 0$, the MOP is optimal for both speech distortion and overall noise reduction, i.e., Pareto-2(0). As λ_x decreases, the relative importance of the ONR term becomes larger, hence a higher SD and a lower ONR are obtained. The scaling parameter η_v controls the amount of overall noise reduction required in the ONR MSE term, such that the output of the filters is a sum of the output signals of the Pareto-2($\eta_v = 0$), weighted with $(1 - \eta_v)$, and the noisy reference microphone signals, weighted with η_v. As η_v increases, a larger component of the noisy reference microphone signals leaks to the output. In the following, it is shown that this binaural beamformer is equivalent to the MWF-N proposed in [7].

5) Signal-Based and TF-Based Pareto BMWF Variants: In this section, variants of the Pareto-2 and Pareto-3 MOPs are introduced. A variant of the generalized cost function for the Pareto-2 MOP is given by substituting $J_x(\boldsymbol{w})$ with $T_x(\boldsymbol{w})$ in (43), i.e.,

$$J_{\text{Pareto-2,b}(\eta_v)}(\boldsymbol{w}) = \lambda_{x,b} T_x(\boldsymbol{w}) + (1 - \lambda_{x,b}) J_{v,\eta_v}(\boldsymbol{w}). \tag{59}$$

Since the signal-based cost function $J_x(\boldsymbol{w})$ is equivalent to the TF-based cost function $T_x(\boldsymbol{w})$, up to a scaling factor, which is equal to the PSD of the signal source (20), we postulate that the two variants of the Pareto-BMWF are equivalent, up to a scaling factor between the weighting parameters such that $\lambda_{x,b} = P_X \lambda_x$, provided that the signals are non-zero.[3]

Similarly, three variants of the proposed Pareto-3 MOP can be obtained by minimizing TF-based cost functions $T_x(\boldsymbol{w})$ and $T_{u,\eta_u}(\boldsymbol{w})$, instead of the signal-based cost functions of the target source and the interfering source $J_x(\boldsymbol{w})$ and $J_{u,\eta_u}(\boldsymbol{w})$, i.e.,

$$\min_{\boldsymbol{w}} \text{Pareto } \mathcal{C}_{\text{Pareto-3b}}(\boldsymbol{w}) \tag{60}$$

$$\min_{\boldsymbol{w}} \text{Pareto } \mathcal{C}_{\text{Pareto-3c}}(\boldsymbol{w}) \tag{61}$$

$$\min_{\boldsymbol{w}} \text{Pareto } \mathcal{C}_{\text{Pareto-3d}}(\boldsymbol{w}) \tag{62}$$

with

$$\mathcal{C}_{\text{Pareto-3b}}(\boldsymbol{w}) = [T_x(\boldsymbol{w}), J_{u,\eta_u}(\boldsymbol{w}), J_{n,\eta_n}(\boldsymbol{w})], \tag{63}$$

$$\mathcal{C}_{\text{Pareto-3c}}(\boldsymbol{w}) = [J_x(\boldsymbol{w}), T_{u,\eta_u}(\boldsymbol{w}), J_{n,\eta_n}(\boldsymbol{w})], \tag{64}$$

$$\mathcal{C}_{\text{Pareto-3d}}(\boldsymbol{w}) = [T_x(\boldsymbol{w}), T_{u,\eta_u}(\boldsymbol{w}), J_{n,\eta_n}(\boldsymbol{w})]. \tag{65}$$

As a result, the generalized MSE cost functions for these three variants are given by

$$J_{\text{Pareto-3,b}(\eta_u,\eta_n)}(\boldsymbol{w}) =$$
$$\lambda_{x,b} T_x(\boldsymbol{w}) + \lambda_{u,b} J_{u,\eta_u}(\boldsymbol{w}) + \lambda_{n,b} J_{n,\eta_n}(\boldsymbol{w}), \tag{66}$$
$$\text{s.t. } \lambda_{x,b} + \lambda_{u,b} + \lambda_{n,b} = 1.$$

$$J_{\text{Pareto-3,c}(\eta_u,\eta_n)}(\boldsymbol{w}) =$$
$$\lambda_{x,c} J_x(\boldsymbol{w}) + \lambda_{u,c} T_{u,\eta_u}(\boldsymbol{w}) + \lambda_{n,c} J_{n,\eta_n}(\boldsymbol{w}), \tag{67}$$
$$\text{s.t. } \lambda_{x,c} + \lambda_{u,c} + \lambda_{n,c} = 1.$$

$$J_{\text{Pareto-3,d}(\eta_u,\eta_n)}(\boldsymbol{w}) =$$
$$\lambda_{x,d} T_x(\boldsymbol{w}) + \lambda_{u,d} T_{u,\eta_u}(\boldsymbol{w}) + \lambda_{n,d} J_{n,\eta_n}(\boldsymbol{w}),$$
$$\text{s.t. } \lambda_{x,d} + \lambda_{u,d} + \lambda_{n,d} = 1. \tag{68}$$

Using the relation in (20) and (23), we postulate that the four variants of the Pareto-3 MOP (38), (66), (67), and (68) are equivalent, except for a scaling factor between the weighting parameters, provided that the signals are non-zero.

IV. CUE PRESERVATION PROPERTIES

The preservation of the sources' binaural cues plays an important role in speech intelligibility improvement. The proposed

---

[3]The PSD of the signal P_X may be larger than $\frac{1}{\lambda_x}$. Since it is required that $0 < \lambda_{x,b} < 1$, additional constraints on the λ_x may be required.

binaural beamformer allows the binaural cues of the target source and the interfering source to be controlled separately using the weighting and scaling parameters. Since the binaural cues can be computed from the ITF using (11), we examine the ITF preservation of the target and the interference sources.

We note that, in general, the binaural cues of the sources are distorted. The trade-off parameters control the amount of distortion. In this section, we examine the impact of the scaling and weighting parameters on the binaural cue preservation.

A. Pareto-3 Setting With Scaling Parameters Set to Zero

Setting both *scaling parameters* η_u and η_n to zero in (36) results with a reasonable requirement to reject the interference and noise completely. By substituting the rank-1 correlation matrix $\boldsymbol{R}_X$ (8) into (44), we obtain binaural filters that are equal to

$$\boldsymbol{w}_L = \lambda_x \boldsymbol{R}_\lambda^{-1} P_X \boldsymbol{a} a_L^*,$$
$$\boldsymbol{w}_R = \lambda_x \boldsymbol{R}_\lambda^{-1} P_X \boldsymbol{a} a_R^*. \qquad (69)$$

For this case, the output RTF of the target component is equal to the input RTF, i.e.,

$$\mathrm{ITF}_{X,\mathrm{OUT}} = \frac{\boldsymbol{w}_L^H \boldsymbol{x}}{\boldsymbol{w}_R^H \boldsymbol{x}} = \frac{a_L}{a_R} = \mathrm{ITF}_{X,\mathrm{IN}}. \qquad (70)$$

Hence, the binaural cues of the target source are preserved. However, the output RTF of the interfering source is also equal to the input RTF of the target source, since the binaural filters are parallel, i.e., $\boldsymbol{w}_L = \mathrm{ITF}_{X,\mathrm{IN}}^* \boldsymbol{w}_R$, such that all sources are perceived as coming from the target source direction (cf. Section V-A), i.e.,

$$\mathrm{ITF}_{U,\mathrm{OUT}} = \frac{\boldsymbol{w}_L^H \boldsymbol{u}}{\boldsymbol{w}_R^H \boldsymbol{u}} = \mathrm{ITF}_{X,\mathrm{IN}}. \qquad (71)$$

This is obtained for Pareto-3($\eta_u = 0, \eta_n = 0$) MOP, for any (λ_x, λ_u), i.e., for all points on the frontier.

Nevertheless, the filters of the Pareto-BMWF are in general not parallel, thus allowing to separately control the binaural cues of the target and the interfering sources.

B. Pareto-3 Setting With Interference Scaling Parameter Larger Than Zero

Setting the *scaling parameter* η_u to a value larger than zero, while the *scaling parameter* η_n is still set to zero, results in an additional binaural filter that is added to (69), i.e.,

$$\boldsymbol{w}_L = \lambda_x \boldsymbol{R}_\lambda^{-1} P_X \boldsymbol{a} a_L^* + \eta_u \lambda_u \boldsymbol{R}_\lambda^{-1} P_U \boldsymbol{b} b_L^*,$$
$$\boldsymbol{w}_R = \lambda_x \boldsymbol{R}_\lambda^{-1} P_X \boldsymbol{a} a_R^* + \eta_u \lambda_u \boldsymbol{R}_\lambda^{-1} P_U \boldsymbol{b} b_R^*. \qquad (72)$$

For this case, the output RTF of the target source is now equal to

$$\mathrm{ITF}_{X,\mathrm{OUT}} = \frac{\frac{P_X \gamma_a}{P_U \gamma_{ba}} \frac{\lambda_x}{\lambda_u}}{\frac{P_X \gamma_a}{P_U \gamma_{ba}} \frac{\lambda_x}{\lambda_u} + \frac{b_R}{a_R} \eta_u} \mathrm{ITF}_{X,\mathrm{IN}}$$
$$+ \frac{\frac{b_R}{a_R} \eta_u}{\frac{P_X \gamma_a}{P_U \gamma_{ba}} \frac{\lambda_x}{\lambda_u} + \frac{b_R}{a_R} \eta_u} \mathrm{ITF}_{U,\mathrm{IN}}, \qquad (73)$$

and the output RTF of the interfering source is equal to

$$\mathrm{ITF}_{U,\mathrm{OUT}} = \frac{\frac{P_X \gamma_{ab}}{P_U \gamma_b} \frac{\lambda_x}{\lambda_u}}{\frac{P_X \gamma_{ab}}{P_U \gamma_b} \frac{\lambda_x}{\lambda_u} + \frac{b_R}{a_R} \eta_u} \mathrm{ITF}_{X,\mathrm{IN}}$$
$$+ \frac{\frac{b_R}{a_R} \eta_u}{\frac{P_X \gamma_{ab}}{P_U \gamma_b} \frac{\lambda_x}{\lambda_u} + \frac{b_R}{a_R} \eta_u} \mathrm{ITF}_{U,\mathrm{IN}}, \qquad (74)$$

with

$$\gamma_{ab} = \boldsymbol{a}^H \boldsymbol{R}_\lambda^{-1} \boldsymbol{b},$$
$$\gamma_a = \boldsymbol{a}^H \boldsymbol{R}_\lambda^{-1} \boldsymbol{a},$$
$$\gamma_b = \boldsymbol{b}^H \boldsymbol{R}_\lambda^{-1} \boldsymbol{b}, \qquad (75)$$

whereas γ_a and γ_b defined as the generalized squared norms, and γ_{ab} defined as the generalized inner product between the ATF vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, and $\gamma_{ba} = \gamma_{ab}^*$.

The output RTFs of the target and interfering sources are now two different weighted sums of the input RTFs of the target and interfering sources. For $\eta_u = 0$, the output RTFs of both the target and interfering sources are equal to the input RTF of the target source $\mathrm{ITF}_{X,\mathrm{IN}}$. For η_u larger than zero, as λ_x becomes larger than $\lambda_u \eta_u$, the output RTFs become "closer" to $\mathrm{ITF}_{X,\mathrm{IN}}$ and vice versa. The output RTF of the target source $\mathrm{ITF}_{X,\mathrm{OUT}}$ is controlled by the ratio between the generalized squared norm γ_a and the generalized inner product γ_{ba}, whereas the output RTF of the interfering source $\mathrm{ITF}_{U,\mathrm{OUT}}$ is controlled by the ratio between the generalized squared norm γ_b and the generalized inner product γ_{ab}. In addition, consider $\frac{P_X}{P_U}$ is the ratio between the powers of the target and interfering sources. For larger $\frac{P_X}{P_U}$, the output RTFs of both target and interfering sources are "closer" to the input RTF of the target $\mathrm{ITF}_{X,\mathrm{IN}}$ and vice versa.

Let us consider a specific optimal set of solutions in the Pareto-3 (η_u, η_n) frontier, where both $T_x(\boldsymbol{w}) = 0$ and $T_{u,\eta_u}(\boldsymbol{w}) = 0$. Substituting (26) and (27) in (14) (as described in Section (II-C), Remark 3) and comparing the results with (16), it is evident that, both the binaural cues of the target and the interference sources are preserved (for any $\eta_u > 0$ and any η_n), i.e.,

$$\mathrm{ITF}_{X,\mathrm{OUT}} = \frac{\boldsymbol{w}_L^H \boldsymbol{a}}{\boldsymbol{w}_R^H \boldsymbol{a}} = \mathrm{ITF}_{X,\mathrm{IN}}; \quad \mathrm{ITF}_{U,\mathrm{OUT}} = \frac{\boldsymbol{w}_L^H \boldsymbol{b}}{\boldsymbol{w}_R^H \boldsymbol{b}} = \mathrm{ITF}_{U,\mathrm{IN}}. \qquad (76)$$

In the following, it is shown that this binaural beamformer is equivalent to the BLCMV-N proposed in [21].

C. Pareto-2 Setting

If the interfering source and the background noise are treated similarly, i.e., the Pareto-2 setting where $\lambda_u = \lambda_n$ and $\eta_u = \eta_n$, an interesting phenomenon appears in relation to the binaural cues. For this case, the output RTF of the target component is equal to the input RTF for all trade-off parameters η_v, i.e.,[4]

$$\mathrm{ITF}_{X,\mathrm{OUT}} = \frac{(1 - \eta_v) \lambda_x P_X \bar{\gamma}_a a_L + \eta_v a_L}{(1 - \eta_v) \lambda_x P_X \bar{\gamma}_a a_R + \eta_v a_R} = \mathrm{ITF}_{X,\mathrm{IN}}. \qquad (77)$$

---

[4]Note that $\bar{\gamma}_a$ and $\bar{\gamma}_{ab}$ are defined similarly as γ_a and γ_{ab} by substituting $\boldsymbol{R}_\lambda$ with $\bar{\boldsymbol{R}}_\lambda$ in (75), respectively.

TABLE I
CONFIGURATION SETUP FOR PARETO-BMWF

	MOP	Target	Interference	Background Noise	Overall Noise	η_u	η_n	η_v	Ref
MWF	Pareto-2	$J_x(\boldsymbol{w})$; Sig-based			$J_{v,0}(\boldsymbol{w})$;Sig-based	-	-	0	[8]
MWF-N	Pareto-2	$J_x(\boldsymbol{w})$; Sig-based			$J_{v,\eta_v}(\boldsymbol{w})$;Sig-based	-	-	η_v	[7]
MWF-IR	Pareto-3	$J_x(\boldsymbol{w})$; Sig-based	$H_{u,\eta_u}(\boldsymbol{w})$;TF-based	$J_{n,0}(\boldsymbol{w})$;Sig-based		η_u	0	-	[40]
BMVDR	Pareto-2	$H_x(\boldsymbol{w})$; TF-based			$J_{v,0}(\boldsymbol{w})$;Sig-based	-	-	0	[14]
BLCMV	Pareto-3	$H_x(\boldsymbol{w})$; TF-based	$H_{u,\eta_u}(\boldsymbol{w})$;TF-based	$J_{n,0}(\boldsymbol{w})$;Sig-based		η_u	0	-	[12]
BMVDR-N	Pareto-2	$H_x(\boldsymbol{w})$; TF-based			$J_{v,\eta_v}(\boldsymbol{w})$;Sig-based	-	-	η_v	[24]
BLCMV-N	Pareto-3	$H_x(\boldsymbol{w})$; TF-based	$H_{u,\eta_u}(\boldsymbol{w})$;TF-based	$J_{n,\eta_n}(\boldsymbol{w})$;Sig-based		η_u	η_n	-	[21]

By substituting (57) in (14), it can be shown that the output RTF of the interfering source is equal to (compare with [38])

$$\text{ITF}_{U,\text{OUT}} = \frac{P_X \bar{\gamma}_{ab} \lambda_x (1 - \eta_v)}{P_X \bar{\gamma}_{ab} \lambda_x (1 - \eta_v) + \frac{b_R}{a_R} \eta_v} \text{ITF}_{X,\text{IN}}$$
$$+ \frac{\frac{b_R}{a_R} \eta_v}{P_X \bar{\gamma}_{ab} \lambda_x (1 - \eta_v) + \frac{b_R}{a_R} \eta_v} \text{ITF}_{U,\text{IN}}. \quad (78)$$

Equation (78) shows that the output RTF of the interfering source is a weighted sum of the input RTF of the target and the input RTF of the interfering source.

Both λ_x and η_v control the output RTF of the interfering source. For $\lambda_x \to 0$, a controllable null steering is obtained (i.e., $J_{v,\eta_v}(\boldsymbol{w}) = 0$), such that the output RTF of the interfering source is equal to the input RTF of the interfering source. If $\eta_v = 0$, the output RTF of the interfering source is equal to the input RTF of the *target* (for any value of $\lambda_x > 0$), whereas, if $\eta_v = 1$, the output RTF of the interfering source is equal to the input RTF of the *interfering* source (as the output filters signals are equal to the noisy input reference microphone signals).

V. SPECIAL CASES OF PARETO-BMWF

In this section, we show that the criteria for several well-known binaural beamformers are special cases of the Pareto-BMWF for specific settings of the scaling and weighting parameters (cf. Table 1), i.e., MWF-based beamformers (Section V-A) and MVDR-based beamformers (Section V-B). In addition, we give insight into the relation between the binaural MWF and the binaural MVDR beamformer (Section V-C).

A. Multi-Channel Wiener Filter-Based Beamformers

In this section, we show that the binaural MWF and several MWF-based binaural beamformers aimed at preserving the binaural cues of the target source, the interfering source and/or the background noise are special cases of the proposed Pareto-BMWF.

1) Binaural MWF

The traditional MWF produces the minimum MSE estimate of the target source component and the output signals, i.e.,

$$J_{\text{traditional MWF}}(\boldsymbol{w}) = \mathcal{E}\left\{ \left\| \begin{bmatrix} x_L - \boldsymbol{w}_L^H \boldsymbol{y} \\ x_R - \boldsymbol{w}_R^H \boldsymbol{y} \end{bmatrix} \right\|^2 \right\}. \quad (79)$$

Considering the target and the noise are statistically independent, the cost function can be written as a sum of the SD term in (17)

and the ONR term in (25), i.e.,

$$J_{\text{traditional MWF}}(\boldsymbol{w}) = J_x(\boldsymbol{w}) + J_{v,0}(\boldsymbol{w}), \quad (80)$$

where $J_{v,0}(\boldsymbol{w})$ is an abbreviation of $J_{v,\eta_v}(\boldsymbol{w})$, where $\eta_v = 0$. In [8], [11], and [39] an extension of the traditional MWF, denoted the binaural MWF, minimizes a weighted sum of the SD term in (17) and the ONR term in (25), i.e.,

$$J_{\text{MWF}}(\boldsymbol{w}) = J_x(\boldsymbol{w}) + \mu_x J_{v,0}(\boldsymbol{w}), \quad (81)$$

where μ_x trades off the speech distortion and noise reduction. Note this beamformer is also referred to as speech distortion weighted MWF (SDW-MWF). In [8], [11], it was shown that the binaural MWF preserves the binaural cues of the target source but distorts the binaural cues of the overall noise, i.e., interfering source plus background noise.

In the following, we show that the binaural MWF is a special case of the proposed Pareto-BMWF. In the Pareto-2 cost function (41), we examine a configuration where the interfering source and the background noise are treated similarly, i.e., $\lambda_v \triangleq \lambda_u = \lambda_n$ and we set $\eta_v \triangleq \eta_u = \eta_n = 0$. When we set $\eta_v = 0$, the generalized cost function in (43) is equal to

$$J_{\text{Pareto-2}(\eta_v=0)}(\boldsymbol{w}) = \lambda_x J_x(\boldsymbol{w}) + (1 - \lambda_x) J_{v,0}(\boldsymbol{w}) \quad (82)$$

where $\lambda_x = 1 - \lambda_v$. Furthermore,

$$J_{\text{Pareto-2}(\eta_v=0)}(\boldsymbol{w}) = \lambda_x \underbrace{(J_x(\boldsymbol{w}) + \mu_x J_{v,0}(\boldsymbol{w}))}_{J_{\text{MWF}}}, \quad (83)$$

with $\mu_x = \frac{1 - \lambda_x}{\lambda_x}$. Clearly, (83) is equivalent to the binaural MWF cost function, except for a constant scaling factor (cf. Table 1).

2) Binaural MWF with partial noise estimation (MWF-N)

In [7], [11], an extension of the binaural MWF, denoted as the MWF with partial noise estimation (MWF-N), was introduced, aiming at the preservation of the binaural cues of the overall noise component, while sacrificing the overall noise reduction. The cost function for the MWF-N is equal to [11]

$$J_{\text{MWF-N}}(\boldsymbol{w}) = J_x(\boldsymbol{w}) + \mu_x J_{v,\eta_v}(\boldsymbol{w}) \quad (84)$$

In [8], [11], it was shown that, while the binaural cues of the target source are preserved, there is a trade-off between overall noise reduction and the preservation of the binaural cues of the overall noise component.

In the following, in a way similar to that for the binaural MWF, we show that the MWF-N is a special case of the proposed Pareto-BMWF. We now examine a configuration where the interfering source and the background noise are treated similarly, i.e., $\lambda_v \triangleq \lambda_u = \lambda_n$ and we set $\eta_v \triangleq \eta_u = \eta_n$ (i.e., strictly larger

than zero), i.e., (41). The solution for this problem defines a Pareto frontier for the Pareto-2(η_v) MOP. Using the scalarization method, the generalized cost function is equal to (43). Furthermore,

$$J_{\text{Pareto-2}(\eta_v)}(\boldsymbol{w}) = \lambda_x \underbrace{(J_x(\boldsymbol{w}) + \mu_x J_{v,\eta_v}(\boldsymbol{w}))}_{J_{\text{MWF-N}}}, \quad (85)$$

with $\mu_x = \frac{1-\lambda_x}{\lambda_x}$. Clearly, (85) is equivalent to the MWF-N cost function, except for a constant scaling factor (cf. Table 1).

3) Binaural MWF with interference reduction (MWF-IR)

It was proposed in [40] that the amount of interference reduction can be controlled using the interference scaling parameter η_u, leading to a novel binaural beamformer denoted as the MWF with interference reduction (MWF-IR). The cost function of the MWF-IR can be written as

$$J_{\text{MWF-IR}}(\boldsymbol{w}) = J_x(\boldsymbol{w}) + \mu_x J_{v,0}(\boldsymbol{w}), \text{ s.t. } \boldsymbol{w} \in H_{u,\eta_u}(\boldsymbol{w}). (86)$$

Since the interfering source component $\boldsymbol{u}$ is constrained by $H_{u,\eta_u}(\boldsymbol{w})$, it can be shown that the solution minimizing (86) is equivalent to the solution minimizing

$$J_{\text{MWF-IR}}(\boldsymbol{w}) = J_x(\boldsymbol{w}) + \mu_x J_{n,0}(\boldsymbol{w}), \text{ s.t. } \boldsymbol{w} \in H_{u,\eta_u}(\boldsymbol{w}), (87)$$

where $J_{v,0}(\boldsymbol{w})$ is substituted by $J_{n,0}(\boldsymbol{w})$.[5] For $\eta_u = 0$, a null is steered toward the interfering source, while the binaural cues of the target source are preserved. For a higher value of η_u, the MWF-IR is able to preserve the binaural cues of the interfering source (due to the hard constraint $H_{u,\eta_u}(\boldsymbol{w})$), whereas the binaural cues of the target source may be distorted.

In the following, we show that the MWF-IR is a special case of the proposed Pareto-BMWF such that the filter solving the MWF-IR cost function lies on the Pareto frontier of the Pareto-3 ($\eta_u, \eta_n = 0$) MOP. Let us examine a configuration, where we set $\eta_n = 0$ for the Pareto-BMWF (61). The MOP is equal to

$$\underset{\boldsymbol{w}}{\text{minPareto }} \mathcal{C}_{\text{Pareto-3c}(\eta_u, \eta_n = 0)}(\boldsymbol{w}) \quad (88)$$

with

$$\mathcal{C}_{\text{Pareto-3c}(\eta_u, \eta_n = 0)}(\boldsymbol{w}) = [J_x(\boldsymbol{w}), T_{u,\eta_u}(\boldsymbol{w}), J_{n,0}(\boldsymbol{w})]. \quad (89)$$

The solution for this problem defines a Pareto frontier for the Pareto-3 ($\eta_u, \eta_n = 0$) MOP. Now, let us examine (88) for $T_{u,\eta_u}(\boldsymbol{w}) = 0$, i.e.,

$$\mathcal{C}_1(\boldsymbol{w}) = [J_x(\boldsymbol{w}), J_{n,0}(\boldsymbol{w})], \text{ s.t. } T_{u,\eta_u}(\boldsymbol{w}) = 0. \quad (90)$$

Since $T_{u,\eta_u}(\boldsymbol{w}) \geq 0$, clearly, any filter that satisfies (90) lies on the Pareto frontier of (88). Using the scalarization method, the generalized cost function for (90) is equal to

$$J_1(\boldsymbol{w}) = \lambda_x J_x(\boldsymbol{w}) + \lambda_n J_{n,0}(\boldsymbol{w}) \text{ s.t. } \boldsymbol{w} \in H_{u,\eta_u}(\boldsymbol{w}), \quad (91)$$

---

[5]Note that (86) and (87) are theoretically equivalent. However, practically, the obtained filters may be different when estimation errors of the correlation matrices $\boldsymbol{R}_V$ and $\boldsymbol{R}_N$ exist.

where $\lambda_n = 1 - \lambda_x$ and $T_{u,\eta_u}(\boldsymbol{w}) = 0$ is substituted with $\boldsymbol{w} \in H_{u,\eta_u}(\boldsymbol{w})$. Furthermore,

$$J_1(\boldsymbol{w}) = \lambda_x \underbrace{(J_x(\boldsymbol{w}) + \mu_x J_{n,0}(\boldsymbol{w})) \text{ s.t. } \boldsymbol{w} \in H_{u,\eta_u}(\boldsymbol{w})}_{J_{\text{MWF-IR}}}, \quad (92)$$

with $\mu_x = \frac{1-\lambda_x}{\lambda_x}$. Clearly, (92) is equivalent to the MWF-IR cost function, except for a constant scaling factor (cf. Table 1).

B. Distortionless Beamformers

In this section, we show that MVDR-based beamformers are also special cases of the proposed Pareto-BMWF. These beamformers are able to extract the target source without distortion and preserve the binaural cues of the target source. Moreover, these beamformers require only an estimate of the RTF vectors of the target source (and the interfering source), whereas the MWF-based beamformers additionally require an estimate of the target source PSD.

1) Binaural MVDR (BMVDR)

The binaural minimum variance distortionless response (BMVDR) beamformer is a binaural extension of the well-known MVDR beamformer [8], [14], [41], reproducing the target source component at both reference microphones without distortion, while minimizing the overall noise power, i.e.,

$$J_{\text{BMVDR}}(\boldsymbol{w}) = J_{v,0}(\boldsymbol{w}), \text{ s.t. } \boldsymbol{w} \in H_x(\boldsymbol{w}). \quad (93)$$

In the following, in a way similar to that for the MWF-based beamformers, we show that the BMVDR is a special case of the proposed Pareto-BMWF, such that the filter that solves the BMVDR cost function lies on the Pareto frontier of the Pareto-2($\eta_v = 0$) MOP defined in (41).

Let us examine the Pareto-2 configuration in (41), where the interfering source and the background noise are treated similarly, $\lambda_v \triangleq \lambda_u = \lambda_n$, and $\eta_v \triangleq \eta_u = \eta_n = 0$. For $J_x(\boldsymbol{w}) = 0$ (or equivalently, $T_x(\boldsymbol{w}) = 0$), the problem in (41) is equal to

$$J_2(\boldsymbol{w}) = J_{v,0}(\boldsymbol{w}), \text{ s.t. } T_x(\boldsymbol{w}) = 0. \quad (94)$$

Since $T_x(\boldsymbol{w}) \geq 0$, clearly, any filter that solves (94) lies on the Pareto frontier of the Pareto-2($\eta_v = 0$) MOP (41). Furthermore, using Section II-C Remark 3, by substituting $T_x(\boldsymbol{w}) = 0$ with $\boldsymbol{w} \in H_x(\boldsymbol{w})$, (94) is equivalent to the BMVDR cost function in (93) (cf. Table 1).

2) Binaural LCMV (BLCMV)

In order to control the amount of interference reduction, an extension of the BMVDR beamformer was proposed in [12], [16], namely, the binaural linearly constrained minimum variance (BLCMV) beamformer. The BLCMV beamformer reproduces the target source component at both reference microphones without distortion, while minimizing the noise power and reducing the interfering source by the interference scaling parameter η_u in both hearing devices. The BLCMV cost function can be written as

$$J_{\text{BLCMV}}(\boldsymbol{w}) = J_{n,0}(\boldsymbol{w}), \text{ s.t. } \boldsymbol{w} \in H_x(\boldsymbol{w}), \ \boldsymbol{w} \in H_{u,\eta_u}(\boldsymbol{w}). \quad (95)$$

In the following, in a way similar to that for the MWF-IR, we show that the BLCMV is a special case of the proposed Pareto-BMWF, such that the filter that solves the BLCMV cost function lies on the Pareto frontier of the Pareto-BMWF ($\eta_u, \eta_n = 0$) MOP. Let us examine the configuration in (62) for $\eta_n = 0$, i.e,

$$\min_{\boldsymbol{w}} \text{Pareto } \mathcal{C}_{\text{Pareto-3d}(\eta_u, \eta_n=0)}(\boldsymbol{w}) \qquad (96)$$

with

$$\mathcal{C}_{\text{Pareto-3d}(\eta_u, \eta_n=0)}(\boldsymbol{w}) = [T_x(\boldsymbol{w}), T_{u,\eta_u}(\boldsymbol{w}), J_{n,0}(\boldsymbol{w})]. \quad (97)$$

For $T_x(\boldsymbol{w}) = 0$ and $T_{u,\eta_u}(\boldsymbol{w}) = 0$, the cost function is equal to

$$J_3(\boldsymbol{w}) = J_{n,0}(\boldsymbol{w}), \text{ s.t. } T_x(\boldsymbol{w}) = 0, T_{u,\eta_u}(\boldsymbol{w}) = 0. \quad (98)$$

Since $T_x(\boldsymbol{w}) \geq 0$ and $T_{u,\eta_u}(\boldsymbol{w}) \geq 0$, clearly, any filter that solves (98) lies on the Pareto frontier of the Pareto-3 ($\eta_u, \eta_n = 0$) MOP (96). Furthermore, using Section II-C Remark 3, by substituting $T_x(\boldsymbol{w}) = 0$ and $T_{u,\eta_u}(\boldsymbol{w}) = 0$ with $\boldsymbol{w} \in H_x(\boldsymbol{w})$ and $\boldsymbol{w} \in H_{u,\eta_u}(\boldsymbol{w})$, respectively, (98) is equivalent to the BLCMV cost function in (95) (cf. Table 1).

For $\eta_u = 0$, it was shown in [13] that a null is steered toward the interfering source, while the binaural cues of the target source are preserved. For a higher value of η_u, the BLCMV beamformer is able to preserve the binaural cues of both the target and the interfering sources, as shown in [12], [16].

3) Binaural MVDR and Binaural LCMV with partial noise estimation (BMVDR-N/BLCMV-N)

Previously in this section, we referred to the MWF-N, which is aimed at preserving the binaural cues of the overall noise component, while sacrificing the overall noise reduction. The disadvantage of this binaural beamformer is that a distorted response for the target source may result. Two extensions of the MWF-N, denoted as the BMVDR with partial noise estimation (BMVDR-N) beamformer and the BLCMV with partial noise estimation (BLCMV-N) beamformer, are proposed, which reproduce the target source component at both reference microphones without distortion [21], [24].

The BMVDR-N cost function can be written as [24]

$$J_{\text{BMVDR-N}}(\boldsymbol{w}) = J_{v,\eta_v}(\boldsymbol{w}), \text{ s.t. } \boldsymbol{w} \in H_x(\boldsymbol{w}), \quad (99)$$

while the BLCMV-N cost function can be written as [21]

$$J_{\text{BLCMV-N}}(\boldsymbol{w}) = J_{v,\eta_v}(\boldsymbol{w}), \text{ s.t. } \boldsymbol{w} \in H_x(\boldsymbol{w}), \boldsymbol{w} \in H_{u,\eta_u}(\boldsymbol{w}). \quad (100)$$

It is straightforward to show that the BMVDR-N is a special case of the proposed Pareto-BMWF(η_v) by substituting $J_{v,0}(\boldsymbol{w})$ with $J_{v,\eta_v}(\boldsymbol{w})$ in the derivation shown for the BMVDR, such that the filter that solves the BMVDR-N cost function lies on the Pareto frontier of the Pareto-2 (η_v) MOP (cf. Table 1). In addition, it is straightforward to show that the BLCMV-N is a special case of the proposed Pareto-BMWF(η_u, η_n) by substituting $J_{n,0}(\boldsymbol{w})$ with $J_{n,\eta_n}(\boldsymbol{w})$ in the derivation shown for the BLCMV, such that the filter that solves the BLCMV-N cost function lies on the Pareto frontier of the Pareto-3(η_u, η_n) MOP (cf. Table 1). These derivations are omitted for the sake of brevity.

C. Insights Into the Relation Between the Binaural MWF and Binaural MVDR Beamformer

Recall that the well-known binaural MWF is controlled by a weighting parameter μ_x that trades off the speech distortion and noise reduction (81). It has been shown that the optimal filters can be decomposed into a (spatial) binaural MVDR beamformer followed by a single-channel (spectral) Wiener filter, i.e.,

$$\boldsymbol{w}_L = \underbrace{\frac{\rho_{MVDR}}{\mu_x + \rho_{MVDR}}}_{w_{L,\text{post}}} \underbrace{\frac{\boldsymbol{R}_V^{-1}\boldsymbol{a}}{\boldsymbol{a}^H \boldsymbol{R}_V^{-1}\boldsymbol{a}} a_L^*}_{\boldsymbol{w}_{L,\text{MVDR}}}$$

$$\boldsymbol{w}_R = \underbrace{\frac{\rho_{MVDR}}{\mu_x + \rho_{MVDR}}}_{w_{R,\text{post}}} \underbrace{\frac{\boldsymbol{R}_V^{-1}\boldsymbol{a}}{\boldsymbol{a}^H \boldsymbol{R}_V^{-1}\boldsymbol{a}} a_R^*}_{\boldsymbol{w}_{R,\text{MVDR}}}, \qquad (101)$$

where $\rho_{MVDR} = P_X \boldsymbol{a}^H \boldsymbol{R}_V^{-1}\boldsymbol{a}$ is the output signal-to-noise ratio (SNR) of the binaural MVDR beamformer [8]. Now, setting $\mu_x = 0$, the obtained binaural MWF filters become equivalent to the binaural MVDR filters $\boldsymbol{w}_{L,\text{MVDR}}$ and $\boldsymbol{w}_{R,\text{MVDR}}$ solving (94) [8], as can be straightforwardly deduced from (101). However, careful examination of the generalized cost function of the binaural MWF (i.e., $J_{\text{MWF}}(\boldsymbol{w}) = J_x(\boldsymbol{w}) + \mu_x J_{v,0}(\boldsymbol{w})$) shows that, if $\mu_x = 0$ (which corresponds with $\lambda_x \to 1$), the generalized cost function is reduced to $J_{\text{MWF}}(\boldsymbol{w}) = J_x(\boldsymbol{w})$, such that it consists of only the SD term. Clearly, this cost function differs from the generalized cost function of the binaural MVDR beamformer (93). Moreover, the cost function of the binaural MWF (81) varies as a function of μ_x, and consequently the value of the cost function is unbounded.

The analyses in this paper clarify these two pitfalls. In Sections V-A and V-B, we proved that the binaural MWF and the binaural MVDR beamformer are two optimal solutions belonging to the same Pareto set of solutions, namely, Pareto-2($\eta_v = 0$). Both SD and ONR terms are non-negative values (cf. Section II-C, Remark 2). Hence, applying the decision-making procedure on the Pareto-2($\eta_v = 0$) frontier, we are free to select any feasible optimal solution. Two such solutions are 1) SD term equal to zero leading to the binaural MVDR, and 2) the binaural MWF, which is continuously controlled by λ_x.

Following, are a few relevant observations that can shed more light on the relations between the two beamformers: 1) a solution satisfying $J_x(\boldsymbol{w}) = 0$ (or equivalently, $T_x(\boldsymbol{w}) = 0$) exists on the Pareto-2($\eta_v = 0$) frontier; 2) for this setting, the filters must satisfy the constraint set (26) (cf. Section II-C, Remark 3); 3) the optimal solution for this setting is equivalent to the binaural MVDR; 4) the parameter setting for that case corresponds with $\lambda_x = 1, \lambda_v = 0$, hence $\mu_x = \frac{1-\lambda_x}{\lambda_x} = 0$. We conclude that substituting the original binaural MWF cost function (81) with generalized cost function (82) is a required step for solving the MOP using the scalarization procedure to clarify the procedure and limit and balance the generalized cost function.

To summarize, under the Pareto formulation we can easily deduce that the MVDR is a proper member in the family of solutions satisfying the same Pareto frontier as the MWF. This observation also shed some light on the relations between these two widely-used beamformers.

VI. SIMULATIONS WITH NOISY SPEECH SIGNALS

In this section, we validate the analytical expressions for the binaural Pareto-2 and Pareto-3 MOPs derived in Section III, and we compare the performance of the considered MOP solutions for various trade-off parameter settings using simulated signals in a noisy and reverberant environment. First, in Section VI-A, the simulation setup and the algorithm parameters are introduced. In Section VI-B, the Pareto-2 MOP derivation is validated demonstrating the single cost function terms for various settings (e.g., the Pareto L-curve is provided). In Section VI-C, the experimental performance evaluation is given, demonstrating the impact of the trade-off parameters on various performance measures, i.e., in terms of binaural SD and the binaural SINR improvement and the target and interference binaural cue preservation capabilities. Similarly, Section VI-D and Section VI-E are dedicated to the Pareto-3 MOP discussing validation and performance evaluation, respectively.

Note that Section VI-B and Section VI-D verify the theoretical MOP derivation using real data, whereas Section VI-C and Section VI-E demonstrate the performance outcomes of the selected settings.

A. Simulation Setup and Algorithm Parameters

In this section, we compare the performance of the considered algorithm using simulated signals in a noisy and reverberant environment using Behind-The-Ear hearing aids From Oldenburg database [31]. Each of the hearing aids is equipped with 2 microphones. Binaural Behind-the-Ear Impulse Responses (BTE-IRs) measured on an artificial head in a cafeteria were used to generate the signal components. The target speech source was located at $-35°$ and a distance of 117.5 cm, while the interfering speech source was located at $0°$ and a distance of 102 cm. Recorded ambient noise from the cafeteria was added to the speech components. The signals were processed at $f_s = 16$ kHz using a weighted overlap-add framework with a block size of 1024 samples and an overlap of 50% between successive blocks. The input signal-to-interference ratio (SIR) with respect to the interference speaker and the SNR with respect to the background noise were set to 10 dB and 0 dB, respectively. For the estimation procedure, three training sections were used. The first training section consisted of a 2 s segment in which none of the speech sources was active. This segment was used to estimate the covariance matrix of the noise component $\boldsymbol{R}_N$. The second training section consisted of a 2.5 s segment in which the target source was active, but the interfering source was inactive. This segment was used to estimate the noisy target source covariance matrix $\boldsymbol{R}_Y$. The third training section consisted of a 2.5 s segment in which the interfering source was active, but the target source was inactive. This segment was used to estimate the overall noise covariance matrix $\boldsymbol{R}_V$. The covariance matrix of the target speech component was estimated as $\boldsymbol{R}_X = \boldsymbol{R}_Y - \boldsymbol{R}_N$, where a rank-1 approximation of $\boldsymbol{R}_X$ was used.

Several performance measures were used for evaluating the performance of the considered algorithm, i.e., the global binaural SD, the global binaural SINR improvement, the global ILD error, and the global ITD error. The global binaural SD is defined
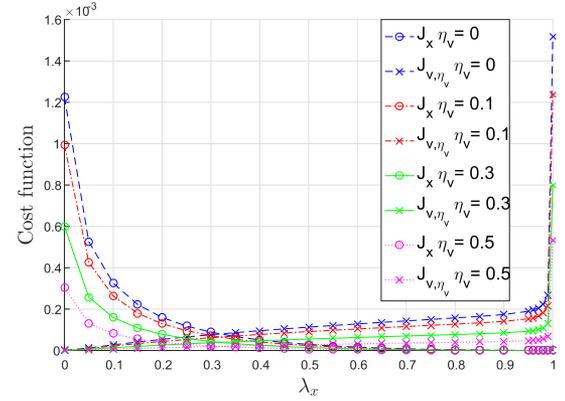
Fig. 1. $J_x(\boldsymbol{w})$ and $J_{v,\eta_v}(\boldsymbol{w})$ as functions of λ_x.

as the narrow-band binaural SD averaged over all frequencies. The global binaural output/input SINR is defined as the average of the narrow-band binaural output/input SINR in dB over all frequencies. The global SINR improvement is obviously defined as the difference between the global binaural output SINR and global binaural input SINR. The narrow-band ILD/ITD error is defined as the absolute value of the difference between the input ILD/ITD and the output ILD/ITD as defined in (11). The global ILD/ITD error is defined as the narrow-band ILD/ITD error averaged over all frequencies.

In the following, we discuss the effects of the weighting and scaling parameters on the binaural SD, binaural SINR, and the binaural cue error for the target and the interfering sources. It is shown that different parameter settings lead to different trade-offs between binaural SD, binaural SINR, and binaural cue errors.

B. Pareto-2 MOP Verification

In the first setting, we examine a configuration where the interfering source and the background noise are treated similarly, i.e., $\lambda_v \triangleq \lambda_u = \lambda_n$, and we set $\eta_v \triangleq \eta_u = \eta_n$ (i.e., higher than zero), i.e.,

$$C_{\text{Pareto-2}(\eta_v)}(\boldsymbol{w}) = \min_{\boldsymbol{w}} \{J_x(\boldsymbol{w}), J_{v,\eta_v}(\boldsymbol{w})\}. \quad (102)$$

Recall the solution for this problem defines a Pareto frontier for the Pareto-2(η_v) MOP such that the generalized cost function is equal to

$$J_{\text{Pareto-2}(\eta_v)}(\boldsymbol{w}) = \lambda_x J_x(\boldsymbol{w}) + (1 - \lambda_x)J_{v,\eta_v}(\boldsymbol{w}), \quad (103)$$

where $\lambda_v = 1 - \lambda_x$. This setting corresponds to the MWF-N (cf. Sec. V-A).

Fig. 1 depicts the cost functions for the SD term $J_x(\boldsymbol{w})$ and for the ONR term $J_{v,\eta_v}(\boldsymbol{w})$, as functions of the target weighting parameter λ_x for various overall noise scaling parameter values η_v. As the target weighting parameter λ_x increases, the cost function for the SD term $J_x(\boldsymbol{w})$ decreases, whereas the cost function for the ONR term $J_{v,\eta_v}(\boldsymbol{w})$ increases. As the overall noise scaling parameter η_v increases, both the cost function for the SD term $J_x(\boldsymbol{w})$ and the cost function for the ONR term

(a) ONR term vs SD term (Pareto L-curve) (b) Zoom of (a)
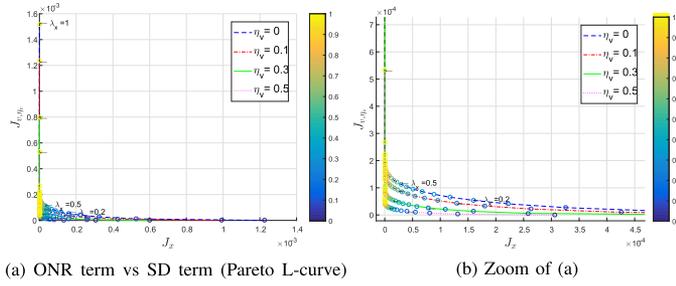
Fig. 2. $J_{v,\eta_v}(\boldsymbol{w})$ as a function of $J_x(\boldsymbol{w})$ for various η_v, color-bar λ_x values.

(a) Output noise PSD $J_{v,0}(\boldsymbol{w})$ and SD $J_x(\boldsymbol{w})$ terms vs λ_x. (b) Output noise PSD term $J_{v,0}(\boldsymbol{w})$ vs SD term $J_x(\boldsymbol{w})$.
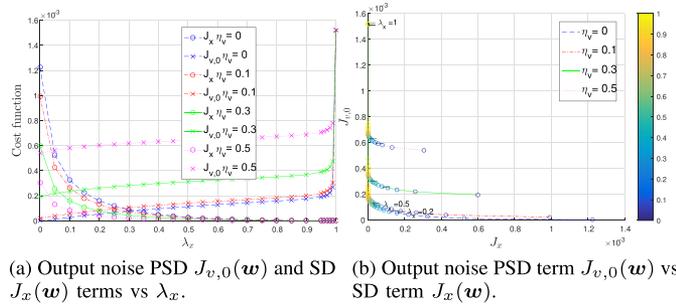
Fig. 3. Noise reduction performance for various η_v: (a) $J_{v,0}(\boldsymbol{w})$ and $J_x(\boldsymbol{w})$ terms as function of λ_x for various η_v; (b) $J_{v,0}(\boldsymbol{w})$ as function of $J_x(\boldsymbol{w})$, color-bar λ_x values.

$J_{v,\eta_v}(\boldsymbol{w})$ decrease, since the MOP requirement is more easily met.

Fig. 2 depicts the cost function for the ONR term $J_{v,\eta_v}(\boldsymbol{w})$ as a function of the cost function for the SD term $J_x(\boldsymbol{w})$, for various target weighting parameter values λ_x and various overall noise scaling parameter values η_v. The Pareto L-curve is clearly seen for various η_v values, where λ_x provides a trade-off between the cost function terms. Clearly, as the overall noise scaling parameter η_v increases, both the cost function for the SD term $J_x(\boldsymbol{w})$ and the cost function for the ONR term $J_{v,\eta_v}(\boldsymbol{w})$ decrease. We emphasize that as shown in Fig. 1-2, as $J_{v,\eta_v}(\boldsymbol{w})$ decreases, $\boldsymbol{w}_L^H \boldsymbol{v}$ and $\boldsymbol{w}_R^H \boldsymbol{v}$ are closer to $\eta_v v_L$ and $\eta_v v_R$, respectively.

C. Performance Evaluation for Pareto 2 MOP

First, we examine the relation between two common performance measures for various settings, i.e., noise reduction versus speech distortion. The amount of noise reduction is evaluated using the output (overall) noise PSD $J_{v,0}(\boldsymbol{w})$, i.e.,

$$J_{v,0}(\boldsymbol{w}) = \mathcal{E}\left\{ \left\| \begin{bmatrix} \boldsymbol{w}_L^H \boldsymbol{v} \\ \boldsymbol{w}_R^H \boldsymbol{v} \end{bmatrix} \right\|^2 \right\}. \tag{104}$$

The amount of speech distortion is evaluated using the signal-based cost function $J_x(\boldsymbol{w})$ for the target SD term (17). Fig. 3(a) and Fig. 3(b) depict the noise reduction performance by substituting $J_{v,\eta_v}(\boldsymbol{w})$ in Fig. 1 and Fig. 2 with $J_{v,0}(\boldsymbol{w})$, respectively. It is observed that the target weighting parameter λ_x trades off the SD term $J_x(\boldsymbol{w})$ and the output noise PSD term $J_{v,0}(\boldsymbol{w})$ for

(a) Binaural SD (b) SINR improvement
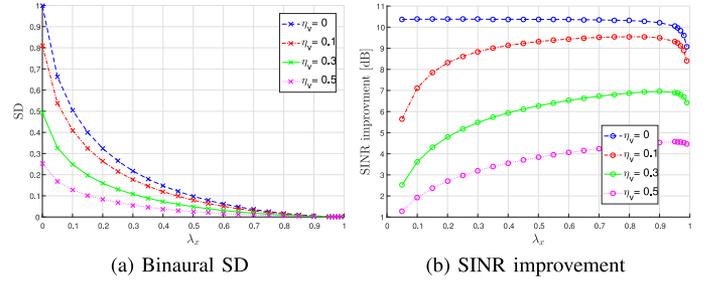
Fig. 4. Dependence of binaural speech distortion (a) and binaural signal-interference-noise ratio improvement (b) on target weighting parameter λ_x.

all various overall noise scaling parameter values η_v such that, as the relative importance of the SD term is higher, the target weighting parameter λ_x increases, and the cost function for the SD term $J_x(\boldsymbol{w})$ decreases, while the output noise PSD term $J_{v,0}(\boldsymbol{w})$ increases.

As the overall noise scaling parameter η_v increases, while the cost function for the SD term $J_x(\boldsymbol{w})$ decreases, the output noise PSD term $J_{v,0}(\boldsymbol{w})$ increases. Interestingly, as $\lambda_x \to 1$, $J_x(\boldsymbol{w}) \to 0$ such that no speech distortion is obtained, whereas as λ_x decreases, the output noise PSD term $J_{v,0}(\boldsymbol{w})$ decreases up to a minimum value (higher than zero for $\eta_v > 0$) obtained for $\lambda_x \to 0$. The minimum value increases as η_v increases (e.g., the L-curve in Fig. 3 is higher). In addition, for this setting (i.e., $\lambda_x \to 0$), $J_x(\boldsymbol{w})$ decreases as η_v increases (e.g., the L-curve in Fig. 3(a) is shorter).

Fig. 4(a) and Fig. 4(b) depict the binaural SD and the binaural SINR improvement performance measures, respectively, as a function of the target weighting parameter λ_x for various overall noise scaling parameter values η_v. As the target weighting parameter λ_x increases, the relative importance of the SD term is higher such that a lower binaural SD is obtained. For $\eta_v = 0$, as the target weighting parameter λ_x increases, the SINR improvement decreases. However, for $\eta_v \neq 0$, this trend does not hold, since λ_x controls the importance of $J_{v,\eta}(\boldsymbol{w})$ rather than of $J_{v,0}(\boldsymbol{w})$. In general, as the overall noise scaling parameter η_v increases, both the binaural SD and the SINR improvement decrease, since the MOP requirement can be more easily met.

Fig. 5 depicts the target and interference binaural cue errors as functions of the target weighting parameter λ_x for various overall noise scaling parameter values η_v. The binaural cues of the target are preserved for any λ_x values, as expected (cf. (77)). As the target weighting parameter λ_x increases, the interference binaural cue errors increase, while as the overall noise scaling parameter η_v increases, the interference binaural cue errors decrease. These trends correspond to Eq. (78). Note that, as the target weighting parameter $\lambda_x \to 1$, the cost function for the SD term $J_x(\boldsymbol{w})$ goes to zero, such that the BMVDR-N is obtained (cf. Sec. V-B).

D. Pareto-3 MOP Verification

In the second setting, we examine a more generalized configuration. The objective now is to minimize simultaneously the

(a) Target ILD error

(b) Target ITD error

(c) Interference ILD error

(d) Interference ITD error
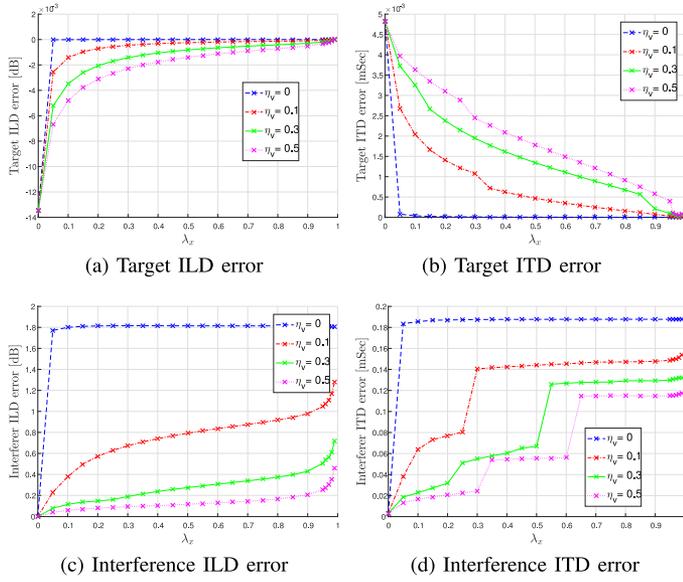
Fig. 5.   Dependence of target binaural cues [ILD (a) and ITD (b)] and interference binaural cues [ILD (c) and ITD (d)] on target weighting parameter λ_x. Note the y-axis value for the target binaural cues is very small.

(a) Cost function vs λ_x

(b) NR term $J_{n,0}(\boldsymbol{w})$ as function of SD term $J_x(\boldsymbol{w})$ for various λ_u
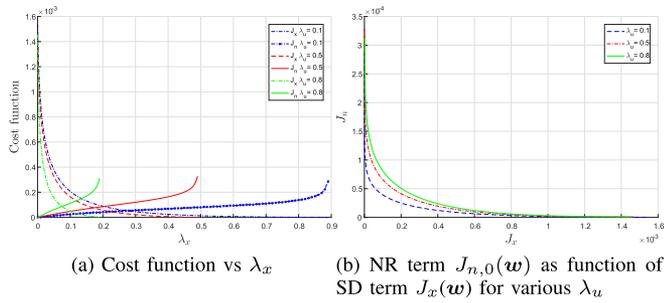
Fig. 6.   Mean squared error cost functions $J_x(\boldsymbol{w})$ and $J_{n,0}(\boldsymbol{w})$ as function of λ_x for various λ_u.

three cost functions, i.e.,

$$\min_{\boldsymbol{w}} \text{Pareto } \mathcal{C}_{\text{Pareto-3}}(\boldsymbol{w}) \qquad (105)$$

with

$$\mathcal{C}_{\text{Pareto-3}}(\boldsymbol{w}) = [J_x(\boldsymbol{w}), J_{u,\eta_u}(\boldsymbol{w}), J_{n,\eta_n}(\boldsymbol{w})], \qquad (106)$$

such that the generalized MSE cost function consists of a weighted sum of the three cost functions (17), (21), and (24), i.e.,

$$J_{\text{Pareto-3}(\eta_u,\eta_n)}(\boldsymbol{w}) = \lambda_x J_x(\boldsymbol{w}) + \lambda_u J_{u,\eta_u}(\boldsymbol{w}) + \lambda_n J_{n,\eta_n}(\boldsymbol{w})$$

$$\text{s.t. } \lambda_x + \lambda_u + \lambda_n = 1. \qquad (107)$$

For this setting the background noise scaling parameter η_n is set to zero, in order to emphasize the noise reduction task, while the interference scaling parameter η_u is set to 0.1. The performance measures are now examined as functions of target weighting parameter λ_x for various values of interference weighting parameter λ_u.

Fig. 6(a) depicts the cost functions for the SD term $J_x(\boldsymbol{w})$ and the NR term $J_{n,0}(\boldsymbol{w})$, as functions of the target weighting

(a) Binaural SD
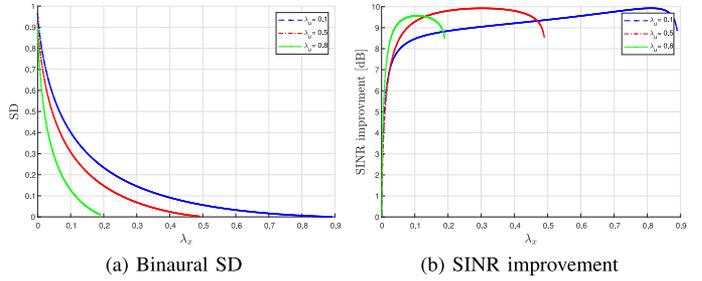
(b) SINR improvement

Fig. 7.   Dependence of binaural SD (a) and binaural signal-interference-noise ratio improvement (b) on target weighting parameter λ_x for various interference weighting parameter values λ_u.

(a) Target ILD error

(b) Target ITD error

(c) Interference ILD error

(d) Interference ITD error
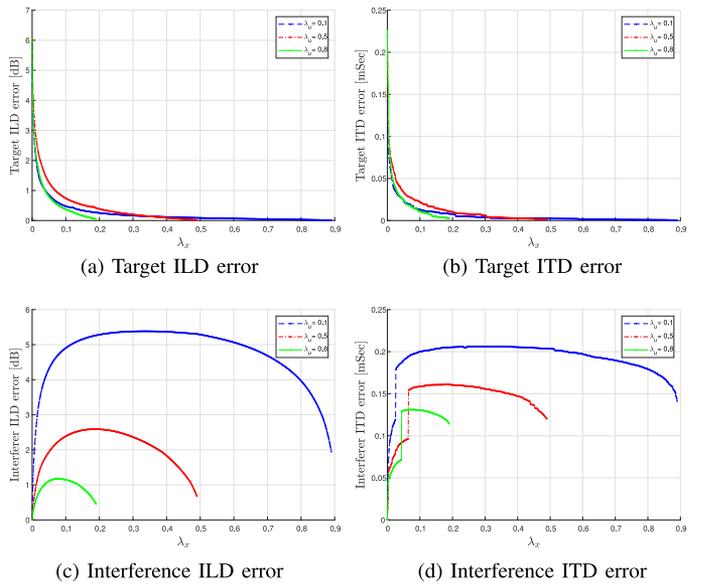
Fig. 8.   Dependence of target binaural cues [ILD (a) and ITD (b)] and interference binaural cues [ILD (c) and ITD (d)] on target weighting parameter λ_x for various interference weighting parameter values λ_u.

parameter λ_x for various interference   weighting parameter values λ_u. Fig. 6(b) depicts the cost function for the NR term $J_{n,0}(\boldsymbol{w})$ as a function of the cost function for the SD term $J_x(\boldsymbol{w})$, for various target weighting parameter λ_x and for various interference weighting parameter values λ_u. As expected, as the target weighting parameter λ_x increases, the cost function for the SD term $J_x(\boldsymbol{w})$ decreases, while the cost function for the NR term $J_{n,0}(\boldsymbol{w})$ decreases. In addition, as the interference weighting parameter λ_u increases, the cost function for the SD term $J_x(\boldsymbol{w})$ decreases, while the cost function for the NR term $J_{n,0}(\boldsymbol{w})$ decreases. This indicates that, as λ_u increases for any specific λ_x, λ_n decreases such that the relative importance of the SD term $J_x(\boldsymbol{w})$ is higher than that of the NR term $J_{n,0}(\boldsymbol{w})$. The Pareto L-curve is clearly observed in Fig. 6(b) for various various λ_u values, where λ_x provides a trade-off between the cost function terms.

E. Performance Evaluation for Pareto 3 MOP

Fig. 7 depicts binaural SD and the binaural SINR improvement performance measures as functions of the target weighting

parameter λ_x for various interference weighting parameter λ_u. As the target weighting parameter λ_x increases, the relative importance of the SD term is higher such that the binaural SD decreases. The same SINR trends as obtained in Fig. 4b for the Pareto-2 case, can also be observed in Fig. 7.

Fig. 8 depicts the target and interference binaural cues as functions of the target weighting parameter λ_x for various interference weighting parameter λ_u. As the target weighting parameter λ_x increases, clearly the target binaural cue errors decrease. However, up to the middle of the range, the interference binaural cue errors increase, while from the middle to the end of the range of the target weighting parameter values the interference binaural cue errors decrease. As the interference weighting parameter λ_u increrases, as expected, the interference binaural cues errors decrease. The interference binaural cue errors are limited with a maximum value that depends on the weighting parameter value λ_u. As the interference weighting parameter λ_u increases, the interference binaural cue errors decrease (this phenomenon corresponds to that mentioned in Section III-D-2).

VII. DISCUSSION AND CONCLUSION

In this paper, we proposed a unified Pareto optimization framework for multi-microphone speech enhancement in binaural hearing aid applications, by defining a generalized MSE cost function, derived from a MOP. As we focused on the dual source scenario, the discussion was restricted to the case of a single target source and a single interfering source, with the corresponding cost functions. We stress, however, that this discussion can be easily extended to cover the multi-target and multi-interference case by introducing additional cost functions to the respective Pareto MOP. An analysis of a specific multi-speaker scenario is provided in [16].

Multiple cost functions can be introduced into the MOP framework. Specifically, an explicit cost function for preserving the binaural cues (e.g., the ITFs of the target and the interfering sources) can be introduced. In the current contribution, we focus only on cost functions addressing signal distortion, interference suppression, and noise reduction, leading to Pareto optimal binaural beamformers. Consequently, the interaural cue preservation can only be implicitly achieved. Nevertheless, we carefully analyzed the binaural cue preservation capabilities of the obtained beamformers. We note that while the binaural cue preservation of the target and interference sources are not an explicit part of the optimization procedure, they are heavily impacted by the application of the optimal beamformers. Three beamformers with explicit binaural cue preservation constraints were presented in [11], [42] and [14]. It can be shown that these beamformers fit the MOP framework and are Pareto optimal.

Two sets of trade-off parameters are provided. The first set, denoted the *scaling* parameters set, η_v and (η_u, η_n), is responsible for the level of the overall, interference, and noise reduction in the respective ONR, IR, and NR cost functions. The Pareto-2(η_v) and Pareto-3(η_u, η_n) frontiers are directly determined by these scaling parameters. The scaling parameters

can be set by the user based on perceptual preferences without resorting to a tedious optimization procedure of a specific performance measure. The scaling parameters may differ in the left and right beamformers and can also be frequency-dependent. In the current contribution, we used, for simplicity, identical values for all frequencies and devices.

The second set of parameters, denoted the *weighting* parameters (λ_x, λ_u), play a different role in the optimization process. These parameters are used to select a specific beamformer that lies on the the Pareto-2(η_v) and Pareto-3(η_u, η_n) frontiers, thus providing the desired trade-off between the SD and ONR, or the SD, IR, and NR cost functions.

The user may use Fig. 8 to determine the weighting parameters that satisfy the permissible level of binaural cue distortion. From these values, the obtained binaural SD, and binaural SINR can be evaluated by using Fig. 7. Note that the experimental study in Section VI is based on real data recordings from Oldenburg database [31] and can therefore facilitate practical design of hearing aids.

Finally, another contribution of the paper is the establishment of the mathematical links between the MVDR and the MWF beamformers. We show in Section V-C that both solutions are specific points on the respective Pareto frontier.

REFERENCES

[1] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge, MA, USA: MIT Press, 1997.

[2] K. Kurozumi and K. Ohgushi, "The relationship between the cross-correlation coefficient of two-channel acoustic signals and sound image quality," *J. Acoustical Soc. Amer.*, vol. 74, no. 6, pp. 1726–1733, Dec. 1983.

[3] A. W. Bronkhorst and R. Plomp, "The effect of head-induced interaural time and level differences on speech intelligibility in noise," *J. Acoustical Soc. Amer.*, vol. 83, no. 4, pp. 1508–1516, Apr. 1988.

[4] 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," *J. Acoustical Soc. Amer.*, vol. 115, no. 2, pp. 833–843, Feb. 2004.

[5] Y. Suzuki, S. Tsukui, F. Asano, and R. Nishimura, "New design method of a binaural microphone array using multiple constraints," *IEICE Tran. Fundamentals Elect., Comm. Comp. Sci.*, vol. 82, no. 4, pp. 588–596, 1999.

[6] T. Lotter and P. Vary, "Dual-channel speech enhancement by superdirective beamforming," *EURASIP J. Appl. Sig. Proc.*, vol. 2006, pp. 1–14, 2006.

[7] T. Klasen, T. Van denM. BogaertMoonen, and J. Wouters, "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," *IEEE Tran. Sig. Proc.*, vol. 55, no. 4, pp. 1579–1585, Apr. 2007.

[8] S. Doclo, S. Gannot, M. Moonen, and A. Spriet, "Acoustic beamforming for hearing aid applications," in *Handbook Array Process. Sensor Netw.*, pp. 269–302, 2010.

[9] K. Reindl, Y. Zheng, and W. Kellermann, "Speech enhancement for binaural hearing aids based on blind source separation," in *4th Int. Symp. Commun., Control Signal Proc.*, 2010, pp. 1–6.

[10] A. Kamkar-Parsi and M. Bouchard, "Instantaneous binaural target PSD estimation for hearing aid noise reduction in complex acoustic environments," *IEEE Tran. Instrum. Meas.*, vol. 60, no. 4, pp. 1141–1154, Apr. 2011.

[11] B. Cornelis, S. Doclo, T. Van dan Bogaert, M. Moonen, and J. Wouters, "Theoretical analysis of binaural multimicrophone noise reduction techniques," *IEEE Tran. Audio, Speech, Lang. Proc.*, vol. 18, no. 2, pp. 342–355, Feb. 2010.

[12] E. Hadad, S. Gannot, and S. Doclo, "Binaural linearly constrained minimum variance beamformer for hearing aid applications," in *Proc. Int. Workshop Acoust. Sig. Enhancement*, Aachen, Germany, 2012, pp. 117–120.

[13] E. Hadad, D. Marquardt, S. Doclo, and S. Gannot, "Binaural multi-channel wiener filter with directional interference rejection," in *Proc. IEEE Int. Conf. Acoust., Speech, Sig. Proc.*, Brisbane, Australia, 2015, pp. 644–648.

[14] E. Hadad, D. Marquardt, S. Doclo, and S. Gannot, "Theoretical analysis of binaural transfer function MVDR beamformers with interference cue preservation constraints," *IEEE/ACM Tran. Audio, Speech, Lang. Proc.*, vol. 23, no. 12, pp. 2449–2464, Dec. 2015.

[15] 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 Tran. Audio, Speech, Lang. Proc.*, vol. 23, no. 12, pp. 2384–2397, Dec. 2015.

[16] E. Hadad, S. Doclo, and S. Gannot, "The binaural LCMV beamformer and its performance analysis," *IEEE/ACM Tran. Audio, Speech, Lang. Proc.*, vol. 24, no. 3, pp. 543–558, Mar. 2016.

[17] A. I. Koutrouvelis, R. C. Hendriks, R. Heusdens, and J. Jensen, "Relaxed binaural LCMV beamforming," *IEEE/ACM Tran. Audio, Speech, Lang. Proc.*, vol. 25, no. 1, pp. 137–152, Jan. 2017.

[18] W. Liao, Z. Luo, I. Merks, and T. Zhang, "An effective low complexity binaural beamforming algorithm for hearing aids," in *Proc. IEEE Workshop Applicat. Sig. Proc. Audio Acoust.*, New Paltz NY, USA, 2015, pp. 1–5.

[19] H. As'ad, M. Bouchard, and H. Kamkar-Parsi, "Perceptually motivated binaural beamforming with cues preservation for hearing aids," in *Proc. IEEE Can. Conf. Elect. Comput. Eng.*, 2016, pp. 1–5.

[20] 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. Eur. Signal Process. Conf.*, 2017, pp. 598–602.

[21] N. Gößling, E. Hadad, S. Gannot, and S. Doclo, "Binaural LCMV beamforming with partial noise estimation," *IEEE/ACM Tran. Audio, Speech, Lang. Proc.*, vol. 28, pp. 2942–2955, 2020.

[22] J. Zhang, R. Heusdens, and R. C. Hendriks, "Rate-distributed binaural LCMV beamforming for assistive hearing in wireless acoustic sensor networks," in *Proc. IEEE 10th Sensor Array Multichannel Signal Process. Workshop*, 2018, pp. 460–464.

[23] W. Pu, J. Xiao, T. Zhang, and Z.-Q. Luo, "A penalized inequality-constrained minimum variance beamformer with applications in hearing aids," in *Proc. IEEE Workshop Appl. Signal Process. to Audio Acoust.*, 2017, pp. 175–179.

[24] D. Marquardt and S. Doclo, "Interaural coherence preservation for binaural noise reduction using partial noise estimation and spectral post-filtering," *IEEE/ACM Tran. Audio, Speech, Lang. Proc.*, vol. 26, no. 7, pp. 1261–1274, Jul. 2018.

[25] S. Doclo, S. Gannot, D. Marquardt, and E. Hadad, "Binaural speech processing with application to hearing devices," *Audio Source Separation Speech Enhancement*, 2018, pp. 413–442.

[26] H. As'ad, M. Bouchard, and H. Kamkar-Parsi, "A robust target linearly constrained minimum variance beamformer with spatial cues preservation for binaural hearing aids," *IEEE/ACM Tran. Audio, Speech, Lang. Proc.*, vol. 27, no. 10, pp. 1549–1563, Oct. 2019.

[27] N. Gößling, D. Marquardt, and S. Doclo, "Performance analysis of the extended binaural MVDR beamformer with partial noise estimation," *IEEE/ACM Tran. Audio, Speech, Lang. Proc.*, vol. 29, pp. 462–476, 2021.

[28] S. Y. Low, S. Nordholm, and K. L. Teo, "Use of efficient frontier in microphone arrays," *Elect. Lett.*, vol. 42, no. 20, pp. 1186–1187, 2006.

[29] K. F. C. Yiu, N. Grbic, K.-L. Teo, and S. Nordholm, "A new design method for broadband microphone arrays for speech input in automobiles," *IEEE Signal Process. Lett.*, vol. 9, no. 7, pp. 222–224, Jul. 2002.

[30] T. L. Vincent and W. J. Grantham, *Optimality in Parametric Systems*. Hoboken, NJ, USA: Wiley, 1981.

[31] H. Kayser, S. Ewert, J. Annemüller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, "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.

[32] T. Marler and S. Arora, "Survey of multi-objective optimization methods for engineering," *Struct. Multidisciplinary Optim.*, vol. 26, no. 6, pp. 369–395, 2004.

[33] M. Caramia and P. Dell'Olmo, "Multi-objective optimization," *Multiobjective Manage. Freight Logistics: Increasing Capacity, Serv. Level Saf. Optim. Algorithms*, 2008, pp. 11–36.

[34] K. Miettinen, *Nonlinear Multiobjective Optimization*. Boston, MA, USA: Kluwer Academic Publishers, 1998.

[35] S. Ruzika and M. M. Wiecek, "Approximation methods in multiobjective programming," *J. Optim. theory Appl.*, vol. 126, no. 3, pp. 473–501, 2005.

[36] G. D. Pelegrina and L. T. Duarte, "A multi-objective approach for blind source extraction," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop*, 2016, pp. 1–5.

[37] T. W. Athan and P. Y. Papalambros, "A note on weighted criteria methods for compromise solutions in multi-objective optimization," *Eng. Optim.*, vol. 27, no. 2, pp. 155–176, 1996.

[38] E. Hadad, D. Marquardt, S. Doclo, and S. Gannot, "Comparison of binaural multichannel wiener filters with binaural cue preservation of the interferer," in *Proc. IEEE Int. Conf. Sci. Elect. Eng.*, 2016, pp. 1–5.

[39] M. Souden, J. Benesty, and S. Affes, "On optimal frequency-domain multichannel linear filtering for noise reduction," *IEEE Tran. Audio, Speech, Lang. Proc.*, vol. 18, no. 2, pp. 260–276, Feb. 2010.

[40] E. Hadad, D. Marquardt, S. Doclo, and S. Gannot, "Extensions of the binaural MWF with interference reduction preserving the binaural cues of the interfering source," in *Proc. IEEE Int. Conf. Acoust., Speech, Sig. Proc.*, Shanghai, China, 2016, pp. 241–245.

[41] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE Tran. Acoust., Speech, Signal Proc.*, vol. 5, no. 2, pp. 4–24, Apr. 1988.

[42] E. Hadad, S. Doclo, and S. Gannot, "A generalized binaural MVDR beamformer with interferer relative transfer function preservation," in *Proc. Eur. Signal Process. Conf.*, 2016, pp. 1643–1647.

Elior Hadad received the B.Sc. (*summa cum laude*) and M.Sc. (*cum laude*) degrees in electrical engineering from the Ben-Gurion University of the Negev, Beer-Sheva, Israel, in 2001 and 2007, respectively, and the Ph.D. degree from Bar-Ilan University, Ramat Gan, Israel, in 2017. He is a Postdoctoral Researcher with the Audio Processing laboratory, Faculty of Engineering, Bar-Ilan University. In 2021, he joined the General Motors Corporation, where he leads an audio algorithms development team. His research interests include statistical signal processing and in particular binaural noise reduction algorithms using microphone arrays, and speaker localization and tracking.

Simon Doclo (Senior Member, IEEE) received the M.Sc. degree in electrical engineering and the Ph.D. degree in applied sciences from Katholieke Universiteit Leuven, Leuven, Belgium, in 1997 and 2003, respectively. From 2003 to 2007, he was a Postdoctoral Fellow with KU Leuven and McMaster University, Hamilton, ON, Canada. From 2007 to 2009, he was a Principal Scientist with NXP Semiconductors in Leuven, Belgium. Since 2009, he has been a Full Professor with the University of Oldenburg, Oldenburg, Germany, and Scientific Advisor for the Oldenburg Branch for Hearing, Speech, and Audio Technology of the Fraunhofer Institute for Digital Media Technology IDMT. His research interests include signal processing for acoustical and biomedical applications, more specifically microphone array processing, speech enhancement, active noise control, acoustic sensor networks, and hearing aid processing. Prof. Doclo was the recipient of several best paper awards (International Workshop on Acoustic Echo and Noise Control 2001, EURASIP Signal Processing 2003, IEEE Signal Processing Society 2008, and VDE Information Technology Society 2019). He is a Member of the IEEE Signal Processing Society Technical Committee on Audio and Acoustic Signal Processing, EURASIP Technical Area Committee on Acoustic, Speech, and Music Signal Processing, and EAA Technical Committee on Audio Signal Processing. Since 2021, he is the Senior Area Editor of IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING.

Sven Nordholm (Senior Member) received the M.Sc. EE (Civilingenjör) and Licentiate of Engineering degrees, and the Ph.D. degree in signal processing from Lund University, Lund, Sweden, in 1983, 1989, and 1992, respectively. Since 1999, Nordholm has been a Professor of signal processing, School of Electrical and Computer Engineering, Curtin University, Bentley WA, Australia. Nordholm was the Research leader Wireless program for the Australian Telecommunication CRC from 1999 to 2006 and the Co-research Leader for a NICTA Project from 2006 to 2009. From 2012 to 2013, he was the Head of Department Electrical Engineering, Curtin University.

He is a co-founder of four start-up companies, Sensear, providing voice communication in extreme noise conditions, Nuheara, a hearables company, Hearmore, a hearing aid company, and Ticking Heart, a biomedical devices Company. Nordholm was a Lead Editor for a special issue on assistive listing techniques in *IEEE Signal Processing Magazine* and several EURASIP special issues. He is a former Associate Editor for *Eurasip Advances in Signal Processing*, the Journal of Franklin Institute, and IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING.

His primary research has encompassed the fields of acoustic signal processing and underwater communications. He has written more than 200 papers in refereed journals and conference proceedings. He frequently contributes to book chapters and encyclopedia articles. He is holding seven patents in the area of speech enhancement and microphone arrays.

Sharon Gannot (Fellow, IEEE) received the B.Sc. degree (*summa cum laude*) in electrical engineering from the Technion-Israel Institute of Technology, Haifa, Israel, in 1986, and the M.Sc. (*cum laude*) and Ph.D. degrees in electrical engineering from Tel-Aviv University, Tel Aviv, Israel, in 1995 and 2000, respectively. In 2001, he held a Postdoctoral position with the Department of Electrical Engineering, KU Leuven, Leuven, Belgium. From 2002 to 2003, he held a Research and Teaching position with the Faculty of Electrical Engineering, Technion-Israel Institute of Technology, Haifa, Israel. He is currently a Full Professor with the Faculty of Engineering, Bar-Ilan University, Ramat-Gan, Israel, where he is the head of the Acoustic Signal Processing Laboratory and the Data Science Program. He is also working as the Faculty Vice Dean and the Deputy Director of the Bar-Ilan University Data Science Institute. From 2018 to 2019, he was also a part-time Professor with the Technical Faculty of IT and Design, Aalborg University, Aalborg, Denmark.

His research interests include statistical signal processing and machine learning algorithms (including manifold learning and deep learning) with applications to single and multi-microphone speech processing. Specifically, distributed algorithms for ad hoc microphone arrays, speech enhancement, noise reduction and speaker separation, dereverberation, single microphone speech enhancement, and speaker localization and tracking. Dr. Gannot was an Associate Editor for the *EURASIP Journal of Advances in Signal Processing* from 2003 to 2012 and the Editor of several special issues on multi-microphone speech processing of the same journal. He was the Guest Editor for several special issues for *Elsevier Speech Communication* and *Signal Processing journals*. He was also the Lead Guest Editor of a special issue for the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, on Acoustic Source Localization and Tracking in Dynamic Real-life Scenes, in 2019. Dr. Gannot was an Associate Editor for the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING from 2009 to 2013 and a Senior Area Chair of the same journal from 2013 to 2017. Since 2020, he has also been a Member of the Senior Editorial Board of the *IEEE Signal Processing Magazine*. He was a Moderator for arXiv in the field of audio and speech processing from 2018 to 2021. He is also a reviewer for many IEEE journals and conferences.

Since January 2010, Dr. Gannot has been a Member of the Audio and Acoustic Signal Processing (AASP) Technical Committee of the IEEE Signal Processing Society (SPS) and was the Committee Chair from 2017 to 2018. He is currently the Chair of the Data Science Initiative of IEEE SPS. Dr. Gannot was a Member of the IEEE SPS, Conferences Board from 2019 to 2021 and a Member of its Executive Subcommittee from 2020 to 2021. Since 2020, he is a Member of the IEEE SPS Education Board and an AE of the newly established Education Center, since 2022. He was an Associate Member of the Special Area Team (SAT) on Acoustic, Speech, and Music Signal Processing from 2016 to 2021. Since 2022, he is the SAT on Signal Processing for Multisensor Processing of the European Association on Signal Processing (EURASIP) and a Member of the European Acoustics Association (EAA), Audio Signal Processing Technical Committee, since 2018. Since 2005, Dr. Gannot has been a Member of the technical and steering committee of the International Workshop on Acoustic Signal Enhancement (IWAENC) and a Member of the Steering Committee, the Latent Variable Analysis and Signal Separation (LVA/ICA), since October 2019. He was the General Co-Chair of IWAENC held in 2010 in Tel-Aviv, Israel, and was the General Co-Chair of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), 2013. He was selected (with colleagues) to present tutorial sessions at ICASSP 2012, EUSIPCO 2012, ICASSP 2013, EUSIPCO 2013, and EUSIPCO 2019, and was a Keynote Speaker for IWAENC 2012, International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA) 2017, Informationstechnische Gesellschaft im VDE (ITG) Conference on Speech Communication 2018, and the Audio Analysis Workshop, 2018.

He was the recipient of Bar-Ilan University Outstanding Lecturer Award, in 2010 and 2014, and the Rector Innovation in Research Award in 2018. He was also the co-recipient of thirteen best paper awards, The EURASIP Group Technical Achievement Award for contributions to theory and practice of microphone array signal processing and statistical learning in speech enhancement through extensive activities of his research group. Dr. Gannot is an IEEE Fellow for contributions to acoustical modelling and statistical learning in speech enhancement (Class 2021).