KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	J. Szurley, A. Bertrand, B. Van Dijk and M. Moonen (2016) **Binaural Noise Cue Preservation in a Binaural Noise Reduction System With a Remote Microphone Signal** IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 24, no. 5, pp. 952-966, May 2016.
Archived version	Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher
Published version	http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7420636
Journal homepage	http://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=6570655
Author contact	joseph.szurley@esat.kuleuven.be *+ 32 (0)16 321795*
IR	https://lirias.kuleuven.be/handle/123456789/532645

(article begins on next page)

Binaural Noise Cue Preservation in a Binaural Noise Reduction System with a Remote Microphone Signal

Joseph Szurley, Alexander Bertrand, *Member, IEEE,* Bas Van Dijk, and Marc Moonen, *Fellow, IEEE*

***Abstract*—A general binaural noise reduction system is considered that employs the multi-channel Wiener filter with partial noise estimation (MWF$_\eta$) allowing for an explicit trade-off between noise reduction and binaural noise cue preservation. In this paper, it is assumed that along with the general binaural system, a remote microphone signal with a high input signal-to-noise ratio (SNR) is available for inclusion in the MWF$_\eta$. The use of this remote microphone signal with a high input SNR allows for a simultaneous increase in both noise reduction performance and preservation of the binaural noise cues. To further increase the performance, a modification to the partial noise estimation (PNE) variable, η, is proposed which relies on exploiting the aforementioned trade-off by either constraining the output SNR or binaural noise cues to the same level before and after the addition of the remote microphone signal. The validity of the theoretical results are supplemented via simulations using a binaural setup with a single speech and noise source.**

***Index Terms*—Binaural hearing aids, binaural cues, noise reduction, multi-channel Wiener filtering**

I. Introduction

HEARING impaired (HI) individuals exhibit poor speech intelligibility in the presence of background noise. Depending on the nature and profoundness of the hearing loss as well as the type of background noise HI individuals may need a higher signal-to-noise ratio (SNR), with a difference of up to 15 dB to have the same intelligibility as non-HI individuals [1]–[3]. In order to reduce the amount of background noise, various auditory prostheses (APs) such as hearing aids (HAs) or cochlear implants (CIs), which come equipped with one or more microphones, perform a noise reduction operation on the microphone signals to provide the individual with an enhanced signal.

Furthermore, for individuals who exhibit hearing impairment in both ears, this can be extended to a bilateral system where two devices are used, e.g., two HAs, two CIs or a bi-modal system, which is a combination of APs. These bilateral APs can operate independently of one another, i.e., *bilateral* processing, where the noise reduction operation is performed only on their own microphone signals. However, an improvement in speech perception in noisy environments has been observed if the APs share microphone signals with one another to perform *binaural* processing [4], [5].

Another advantageous aspect of binaural processing is that it allows for preservation of the so-called binaural cues, namely the interaural level difference (ILD), and the interaural time difference (ITD) or interaural phase difference (IPD), which are both known to contribute to sound localization by the human auditory system [6], [7]. In fact, the ILD and IPD have been shown to be related to the amplitude and phase of the interaural transfer function (ITF) [8]–[10], which describes the transfer function between the left and right ears.

Due to the limited size of the APs, which restricts the spatial diversity of the microphones, as well as challenging noise environments, there has been an increased interest in incorporating remote microphone signals to the APs [11]–[13]. It is assumed that these remote microphones are either worn or placed close to the desired speech source thereby providing an improved SNR resulting in increased speech intelligibility. For example, in classroom and auditorium settings or if the HI individual interacts routinely with the same person(s), e.g., child parent interaction, the microphone can be worn by the speaker providing a higher SNR signal.

There are currently various systems that are able to transmit a remote microphone signal to the APs, e.g., traditional telecoil or FM transmission systems [14], [15] and new state of the art digital wireless transmission systems [16]. However, these systems either disable the local microphones or significantly amplify the remote microphone signal to provide the HI individuals with a clean, high SNR signal [16]. Unfortunately, by disabling the local microphones, the binaural cues become degraded, or in the worst case, completely lost.

We therefore envisage a general binaural system, where each device performs noise reduction with the local microphone

This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of KU Leuven Research Council CoE PFV/10/002 'Optimization in Engineering' (OPTEC), Research Project FWO nr. G.0763.12 'Wireless acoustic sensor networks for extended auditory communication', Research Project FWO nr. G.0931.14 'Design of distributed signal processing algorithms and scalable hardware platforms for energy-vs-performance adaptive wireless acoustic sensor networks, Project BOF/STG-14-005 Research Fund KU Leuven, IWT O&O Project nr. 110722 'Signal processing and automatic fitting for next generation cochlear implants' and HANDiCAMS. The project HANDiCAMS acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number: 323944. The scientific responsibility is assumed by its authors.

KU Leuven, Department of Electrical Engineering (ESAT), Stadius Center for Dynamical Systems, Signal Processing and Data Analytics (A. Bertrand is also with iMinds Medical IT) Kasteelpark Arenberg 10, box 2446, 3001 Leuven, Belgium (e-mail: joseph.szurley@esat.kuleuven.be, alexander.bertrand@esat.kuleuven.be, marc.moonen@esat.kuleuven.be).

B. Van Dijk is with Cochlear CTCE, Schalienhoevedreef 20 I, 2800 Mechelen, Belgium

signals as well as a remote microphone signal[1]. The devices perform this noise reduction by employing the multi-channel Wiener filter (MWF). However, it has been shown that while the MWF perfectly preserves the binaural speech cues, the noise binaural cues are distorted [8]. This distortion may place the listener in danger in situations where the binaural noise cues may be important for the safety, e.g. hazard warnings, traffic, etc.

In order to preserve the binaural speech and noise cues, several methods have been proposed by either conditioning the microphone signals or modifying the MWF. Binaural cue preservation was studied in [17] which relies on applying a gain to the signal, albeit only to preserve the binaural speech cues. In [18] an extension to the MWF was proposed that partially preserves only the noise IPD cues by incorporating a trade-off parameter to the cost function of the MWF. This trade-off parameter allows for better noise IPD cue preservation at the expense of less noise reduction. In [8], [9], this was further extended to a trade-off between several cost functions, namely the MWF, speech ITF and noise ITF. While there is no closed-form expression for the combined noise reduction and noise ITF preservation, in [10] a simplification to the ITF cost function was proposed allowing for a tractable solution. Furthermore in [10], [19] an extension to the MWF was introduced, the MWF with partial noise estimation (MWF$_\eta$), which allows for an explicit trade-off between the amount of noise reduction and the preservation of the binaural noise cues in a closed-form solution.

It will be shown that when a high input SNR remote microphone signal is available to the binaural system, not only is the noise reduction performance increased, but also a better preservation of the binaural noise cues is achieved. To further increase the performance, a modification to the partial noise estimation (PNE) variable is proposed which relies on exploiting the aforementioned trade-off, either constraining the output SNR or binaural noise cues to the same level before and after the addition of the remote microphone signal.

Also when the noise and speech signal come from the same direction, due to the spatial filtering of the MWF$_\eta$, almost no SNR improvement is observed [18], [20]. The addition of the remote microphone can then help to improve the SNR, as it adds more spatial diversity in the signal subspace.

This paper is organized as follows. Section II introduces the general binaural noise reduction system along with the MWF$_\eta$. In Section III the performance measures are defined emphasizing the distortion in the binaural noise cues at the output of the system. In Section IV the general binaural system is extended to include a remote microphone signal and in Section V the effect on the output SNR and binaural noise cues is derived. In Section VI a modification to the PNE variable η is given in order to further increase noise reduction performance or binaural noise cue preservation. Simulations are performed in Section VII, using head-related impulse responses (HRIRs) to generate various acoustic environments where a remote microphone signal is available for inclusion into the MWF$_\eta$. Finally conclusions are given in Section VIII.

[1]We note that the analysis in this paper can easily be extended to a so-called wireless acoustic sensor network, where many remote microphone signals are available [13].

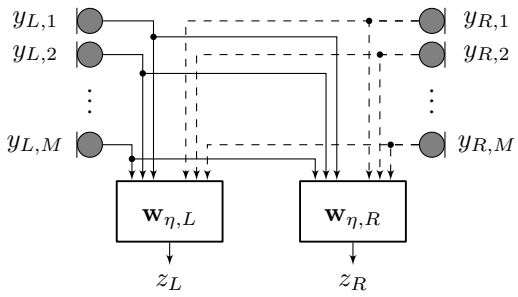


Fig. 1: A general binaural noise reduction system.

II. Data Model

We consider a general binaural noise reduction system as depicted in Figure 1 that consists of two devices which are each equipped with M microphones, and so the combined number of microphones (left and right device) is $2M$. The signal of the mth microphone of the left device, denoted with the subscript $_{L,m}$, is given in the frequency domain as

$$y_{L,m}(\omega) = x_{L,m}(\omega) + n_{L,m}(\omega) \tag{1}$$

where $x_{L,m}(\omega)$ is the speech component, $n_{L,m}(\omega)$ is the noise component, and ω is the discrete frequency variable. The microphones of the right device are represented in the same fashion with R instead of L in the subscript. For the sake of brevity we will omit ω from the following derivations bearing in mind that the operations take place in the frequency domain.

We define a stacked M-dimensional vector containing the microphone signals of the left device as

$$\mathbf{y}_L = \begin{bmatrix} y_{L,1} \\ \vdots \\ y_{L,M} \end{bmatrix} \tag{2}$$

where $\mathbf{x}_L$ and $\mathbf{n}_L$ are defined similarly and contain the speech components and noise components respectively. A stacked M-dimensional vector containing the microphone signals of the right device, $\mathbf{y}_R$, as well as the speech and noise components, $\mathbf{x}_R$ and $\mathbf{n}_R$, are also defined in the same manner as (2). A stacked $2M$-dimensional vector is defined that contains all of the microphone signals of the system and is given as

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_L \\ \mathbf{y}_R \end{bmatrix} \tag{3}$$

where again $\mathbf{x}$ and $\mathbf{n}$ are defined similarly.

Each device estimates the speech component of a reference microphone signal, $x_{\{L,R\},ref}$ where $\{L,R\}$ indicates either L or R, by applying a multi-channel filter $\mathbf{w}_{\eta,\{L,R\}}$ to the microphone signals given in (3). We assume, without loss of generality, that the first microphone is chosen as the reference microphone such that $x_{\{L,R\},ref} = x_{\{L,R\},1}$. Preferably, the reference microphone is chosen as the microphone with highest SNR. In APs, the reference microphone is often

chosen to be the front facing microphone, however, there are other methods that may be used to select the reference microphone that are dependent on the scenario and locations of the microphones [21], [22].

The filtered output of each device is given as

$$z_L = \mathbf{w}_{\eta,L}^H \mathbf{y}, \quad z_R = \mathbf{w}_{\eta,R}^H \mathbf{y} \tag{4}$$

which can be expanded to

$$z_L = \mathbf{w}_{\eta,L}^H \mathbf{x} + \mathbf{w}_{\eta,L}^H \mathbf{n}, \quad z_R = \mathbf{w}_{\eta,R}^H \mathbf{x} + \mathbf{w}_{\eta,R}^H \mathbf{n} \tag{5}$$

where H is the Hermitian transpose.

Each device finds its multi-channel filter based on the linear minimum mean square error criterion between the speech and noise component in its reference microphone signal and the filtered version of the microphone signals [10], i.e.,

$$\begin{aligned}\mathbf{w}_{\eta,L} &= \underset{\mathbf{w}_{\eta,L}}{\arg\min}\ \mathcal{E}\{|x_{L,1} - \mathbf{w}_{\eta,L}^H \mathbf{x}|^2\} + \mu\mathcal{E}\{|\eta n_{L,1} - \mathbf{w}_{\eta,L}^H \mathbf{n}|^2\} \\ \mathbf{w}_{\eta,R} &= \underset{\mathbf{w}_{\eta,R}}{\arg\min}\ \mathcal{E}\{|x_{R,1} - \mathbf{w}_{\eta,R}^H \mathbf{x}|^2\} + \mu\mathcal{E}\{|\eta n_{R,1} - \mathbf{w}_{\eta,R}^H \mathbf{n}|^2\}\end{aligned} \tag{6}$$

where $0 \leq \mu$ is a trade-off parameter emphasizing the noise reduction at the cost of higher speech distortion, $0 \leq \eta \leq 1$ is the PNE variable that allows a certain amount of the noise to be present in the output signal and $\mathcal{E}\{\cdot\}$ is the expectation operator.

The trade-off parameter μ can be chosen based on constraints such as the input SNR [23] or using a perceptual based weighting [24]. Likewise the PNE variable, η, can be chosen based on constraints similar to the method proposed in [25] which tries to preserve a constant noise level, or again a perceptual based weighting [26]. For ease of exposition, we will assume a fixed value of μ and η in the following derivations. Note that η essentially controls how much residual noise is present in the filtered output and can be thought of as producing a given SNR increase of up to a limit of $\frac{1}{\eta^2}$. This will be explained in more detail in Section V-A.

A. General data model

We define a speech and noise correlation matrix as

$$\mathbf{R}_{\mathbf{xx}} = \mathcal{E}\{\mathbf{x}\mathbf{x}^H\},\ \mathbf{R}_{\mathbf{nn}} = \mathcal{E}\{\mathbf{n}\mathbf{n}^H\} \tag{7}$$

and the corresponding input speech and noise powers ($P_{s,\{L,R\}}$ and $P_{n,\{L,R\}}$ respectively) at the reference microphones as

$$P_{s,\{L,R\}} = \mathbf{e}_{\{L,R\}}^T \mathbf{R}_{\mathbf{xx}} \mathbf{e}_{\{L,R\}} \tag{8}$$

$$P_{n,\{L,R\}} = \mathbf{e}_{\{L,R\}}^T \mathbf{R}_{\mathbf{nn}} \mathbf{e}_{\{L,R\}} \tag{9}$$

where $\mathbf{e}_{\{L,R\}}$ are unit vectors with a 1 at the index of the reference microphone signal, $x_{\{L,R\},1}$.

A single speech source is assumed such that $\mathbf{R}_{\mathbf{xx}}$ is a rank-1 matrix that can be represented as

$$\mathbf{R}_{\mathbf{xx}} = P_s \mathbf{a}\mathbf{a}^H \tag{10}$$

where P_s is the speech power and $\mathbf{a}$ is the speech steering vector that describes the transfer function from the speech source to the microphones. This rank-1 assumption is used throughout the derivations in the sequel. If we define the M-dimensional speech steering vectors $\mathbf{a}_{\{L,R\}}$ in the same fashion as (2), then the speech power at the reference microphones (8), can be written as

$$P_{s,\{L,R\}} = P_s |a_{\{L,R\},1}|^2 \,. \tag{11}$$

A voice activity detector (VAD) can be used to compute a correlation matrix $\mathbf{R}_{\mathbf{yy}} = \mathcal{E}\{\mathbf{y}\mathbf{y}^H\}$ during speech-and-noise periods and $\mathbf{R}_{\mathbf{nn}}$ during noise only periods. The speech correlation matrix, $\mathbf{R}_{\mathbf{xx}}$, can be then estimated in several such as simple subtraction, $\mathbf{R}_{\mathbf{xx}} = \mathbf{R}_{\mathbf{yy}} - \mathbf{R}_{\mathbf{nn}}$, or more complex methods which rely on a generalized eigenvalue decomposition of $\mathbf{R}_{\mathbf{yy}}$ and $\mathbf{R}_{\mathbf{nn}}$ [27].

The solution to (6) is the multi-channel Wiener filter with PNE (MWF$_\eta$) which is given as [10]

$$\begin{aligned}\mathbf{w}_{\eta,L} &= (1-\eta)\frac{\mathbf{R}_{\mathbf{nn}}^{-1}\mathbf{R}_{\mathbf{xx}}\mathbf{e}_L}{\mu + \mathrm{Tr}\{\mathbf{R}_{\mathbf{nn}}^{-1}\mathbf{R}_{\mathbf{xx}}\}} + \eta\mathbf{e}_L \\ \mathbf{w}_{\eta,R} &= (1-\eta)\frac{\mathbf{R}_{\mathbf{nn}}^{-1}\mathbf{R}_{\mathbf{xx}}\mathbf{e}_R}{\mu + \mathrm{Tr}\{\mathbf{R}_{\mathbf{nn}}^{-1}\mathbf{R}_{\mathbf{xx}}\}} + \eta\mathbf{e}_R\end{aligned} \tag{12}$$

where $\mathrm{Tr}\{\cdot\}$ is the trace of a matrix. For ease of exposition we represent $\mathrm{Tr}\{\mathbf{R}_{\mathbf{nn}}^{-1}\mathbf{R}_{\mathbf{xx}}\}$ as

$$\lambda \triangleq \mathrm{Tr}\{\mathbf{R}_{\mathbf{nn}}^{-1}\mathbf{R}_{\mathbf{xx}}\} \tag{13}$$

which will be used in the sequel. When $\eta = 0$, (12) corresponds to the MWF with no PNE, whereas $\eta = 1$ corresponds to the trivial case where no noise reduction is performed and the reference microphone signal is passed to the output of the device.

B. Data model with single coherent noise source

If a single coherent (spatially correlated) noise source and additive incoherent noise is present, the noise correlation matrix can given similarly to that of [28] and represented as

$$\mathbf{R}_{\mathbf{nn}} = P_v \mathbf{v}\mathbf{v}^H + \delta\mathbf{I} \tag{14}$$

where P_v is the spatially correlated noise source power, $\mathbf{v}$ is the $2M$-dimensional spatially correlated noise steering vector and δ is the noise power of the incoherent noise. For ease of exposition it will be assumed that noise power is normalized to unity such that $P_v = 1$.

Again, if we define the M-dimensional noise steering vectors $\mathbf{v}_{\{L,R\}}$ in the same fashion as (2), then the input noise power at the reference microphones is given in the same fashion as (11), i.e.,

$$P_{n,\{L,R\}} = |v_{\{L,R\},1}|^2 + \delta \,. \tag{15}$$

Using the matrix inversion lemma, the inverse of (14) is given as

$$\mathbf{R}_{\mathbf{nn}}^{-1} = \frac{1}{\delta}\left[\mathbf{I} - \frac{\mathbf{v}\mathbf{v}^H}{\delta + ||\mathbf{v}||^2}\right] \,. \tag{16}$$

When using (10) and (16) we see that λ can be now be given as

$$\lambda = \frac{P_s}{\delta}\left[||\mathbf{a}||^2 - \frac{|\mathbf{a}^H\mathbf{v}|^2}{\delta + ||\mathbf{v}||^2}\right] \,. \tag{17}$$

We note that while (16) and (17) only hold for a noise correlation matrix as defined in (14), this will be used to motivate some limiting cases of λ. The original definition of λ (see (13)) can be used in all other cases.

III. Performance measures

A. Output SNR

The input SNR for the left and right device using (8) and (9) is given as

$$\text{SNR}_{\text{IN},\{L,R\}} = \frac{\mathbf{e}_{\{L,R\}}^T \mathbf{R}_{\mathbf{xx}} \mathbf{e}_{\{L,R\}}}{\mathbf{e}_{\{L,R\}}^T \mathbf{R}_{\mathbf{nn}} \mathbf{e}_{\{L,R\}}} \tag{18}$$

$$= \frac{P_{s,\{L,R\}}}{P_{n,\{L,R\}}} \tag{19}$$

and when using (11) and (15), can be given as

$$\text{SNR}_{\text{IN},\{L,R\}} = \frac{P_s |a_{\{L,R\},1}|^2}{|v_{\{L,R\},1}|^2 + \delta} \, . \tag{20}$$

The output SNR for the left and right device using (12) is given as

$$\text{SNR}_{\text{OUT},\{L,R\}} = \frac{\mathbf{w}_{\eta,\{L,R\}}^T \mathbf{R}_{\mathbf{xx}} \mathbf{w}_{\eta,\{L,R\}}}{\mathbf{w}_{\eta,\{L,R\}}^T \mathbf{R}_{\mathbf{nn}} \mathbf{w}_{\eta,\{L,R\}}} \tag{21}$$

$$= \frac{\frac{(1-\eta)^2 \lambda^2 P_{s,\{L,R\}}}{(\mu+\lambda)^2} + \frac{2(1-\eta)\eta\lambda P_{s,\{L,R\}}}{(\mu+\lambda)} + \eta^2 P_{s,\{L,R\}}}{\frac{(1-\eta)^2 \lambda P_{s,\{L,R\}}}{(\mu+\lambda)^2} + \frac{2(1-\eta)\eta P_{s,\{L,R\}}}{(\mu+\lambda)} + \eta^2 P_{n,\{L,R\}}} \, . \tag{22}$$

Looking at the limiting cases of η, the output SNR is given as [10], [29]

$$\text{SNR}_{\text{OUT},\{L,R\}} = \begin{cases} \lambda & : \eta = 0 \\ \text{SNR}_{\text{IN},\{L,R\}} & : \eta = 1 \end{cases} \tag{23}$$

where (22) lies in this range for any other value of $0 < \eta < 1$. We also see that for $\eta = 0$, $\text{SNR}_{\text{OUT},L} = \text{SNR}_{\text{OUT},R}$ which are both equal to λ as defined in (13).

As a particular case, we consider the scenario where the noise correlation matrix is given as in (14) and the speech and noise have similar steering vectors, such that $\frac{\mathbf{a}}{\|\mathbf{a}\|} \approx \frac{\mathbf{v}}{\|\mathbf{v}\|}$, for example both emanating from the same direction in free-field conditions. This results in a minimum value for λ in (17) equal to $\frac{P_s \|\mathbf{a}\|^2}{\delta + \|\mathbf{v}\|^2}$. Since λ is the upper limit for the output SNR, a smaller value for λ will produce a smaller overall range of SNR improvement in (23). In the worst case, this can lead to $\text{SNR}_{\text{OUT},\{L,R\}} \approx \text{SNR}_{\text{IN},\{L,R\}}$ as shown experimentally in [9], [20] and in Section VII-D.

B. Binaural cues

We define the speech interaural transfer function (ITF) at the input and output as [9]

$$\text{ITF}_{\text{IN},s} = \frac{\mathbf{e}_L^T \mathbf{R}_{\mathbf{xx}} \mathbf{e}_L}{\mathbf{e}_R^T \mathbf{R}_{\mathbf{xx}} \mathbf{e}_L} = \frac{\mathbf{e}_L^T \mathbf{R}_{\mathbf{xx}} \mathbf{e}_R}{\mathbf{e}_R^T \mathbf{R}_{\mathbf{xx}} \mathbf{e}_R} = \frac{P_{s,L}}{\phi_s} \tag{24}$$

$$\text{ITF}_{\text{OUT},s} = \frac{\mathbf{w}_{\eta,L}^H \mathbf{R}_{\mathbf{xx}} \mathbf{w}_{\eta,L}}{\mathbf{w}_{\eta,R}^H \mathbf{R}_{\mathbf{xx}} \mathbf{w}_{\eta,L}} = \frac{\mathbf{w}_{\eta,L}^H \mathbf{R}_{\mathbf{xx}} \mathbf{w}_{\eta,R}}{\mathbf{w}_{\eta,R}^H \mathbf{R}_{\mathbf{xx}} \mathbf{w}_{\eta,R}} \tag{25}$$

where $\phi_s = \mathbf{e}_R^T \mathbf{R}_{\mathbf{xx}} \mathbf{e}_L$. Substituting (12) into (25) shows that the speech ITF is unaffected by the filtering process, i.e.,

$$\text{ITF}_{\text{IN},s} = \text{ITF}_{\text{OUT},s} \, . \tag{26}$$

The noise ITF is defined in a similar manner at the input and output, i.e.,

$$\text{ITF}_{\text{IN},n} = \frac{\mathbf{e}_L^T \mathbf{R}_{\mathbf{nn}} \mathbf{e}_L}{\mathbf{e}_R^T \mathbf{R}_{\mathbf{nn}} \mathbf{e}_L} = \frac{\mathbf{e}_L^T \mathbf{R}_{\mathbf{nn}} \mathbf{e}_R}{\mathbf{e}_R^T \mathbf{R}_{\mathbf{nn}} \mathbf{e}_R} = \frac{P_{n,L}}{\phi_n} \tag{27}$$

$$\text{ITF}_{\text{OUT},n} = \frac{\mathbf{w}_{\eta,L}^H \mathbf{R}_{\mathbf{nn}} \mathbf{w}_{\eta,L}}{\mathbf{w}_{\eta,R}^H \mathbf{R}_{\mathbf{nn}} \mathbf{w}_{\eta,L}} = \frac{\mathbf{w}_{\eta,L}^H \mathbf{R}_{\mathbf{nn}} \mathbf{w}_{\eta,R}}{\mathbf{w}_{\eta,R}^H \mathbf{R}_{\mathbf{nn}} \mathbf{w}_{\eta,R}} \tag{28}$$

where $\phi_n = \mathbf{e}_R^T \mathbf{R}_{\mathbf{nn}} \mathbf{e}_L$. Substituting (12) into (28) results in

$$\text{ITF}_{\text{OUT},n} = \frac{\frac{(1-\eta)^2 \lambda P_{s,L}}{(\mu+\lambda)^2} + \frac{2(1-\eta)\eta P_{s,L}}{(\mu+\lambda)} + \eta^2 P_{n,L}}{\frac{(1-\eta)^2 \lambda \phi_s}{(\mu+\lambda)^2} + \frac{2(1-\eta)\eta \phi_s}{(\mu+\lambda)} + \eta^2 \phi_n} \tag{29}$$

which means that the $\text{ITF}_{\text{OUT},n}$ will change depending on the value of η. Looking at the limiting cases of η

$$\text{ITF}_{\text{OUT},n} = \begin{cases} \text{ITF}_{\text{OUT},s} = \text{ITF}_{\text{IN},s} & : \eta = 0 \\ \text{ITF}_{\text{IN},n} & : \eta = 1 \end{cases} \tag{30}$$

which means that at lower values of η the noise ITF at the output will be closer to the speech ITF at the input and output whereas at higher values of η the noise ITF at the output is closer to the noise ITF at the input.

Using the ITF, we also define two binaural cues for the speech and noise, the ILD and the IPD, which are related to the amplitude and phase of the ITFs. The speech ILD and IPD at the input are given by

$$\text{ILD}_{\text{IN},s} = |\text{ITF}_{\text{IN},s}| \tag{31}$$

$$= \frac{\mathbf{e}_L^T \mathbf{R}_{\mathbf{xx}} \mathbf{e}_L}{\mathbf{e}_R^T \mathbf{R}_{\mathbf{xx}} \mathbf{e}_R} \tag{32}$$

$$= \frac{P_{s,L}}{P_{s,R}} \tag{33}$$

$$\text{IPD}_{\text{IN},s} = \angle \left(\text{ITF}_{\text{IN},s} \right) \tag{34}$$

$$= \angle \left(\frac{\mathbf{e}_L^T \mathbf{R}_{\mathbf{xx}} \mathbf{e}_R}{\mathbf{e}_R^T \mathbf{R}_{\mathbf{xx}} \mathbf{e}_R} \right) \tag{35}$$

$$= \angle \left(\phi_s \right) \tag{36}$$

and at the output are given by

$$\text{ILD}_{\text{OUT},s} = \frac{\mathbf{w}_{\eta,L}^H \mathbf{R}_{\mathbf{xx}} \mathbf{w}_{\eta,L}}{\mathbf{w}_{\eta,R}^H \mathbf{R}_{\mathbf{xx}} \mathbf{w}_{\eta,R}} \tag{37}$$

$$\text{IPD}_{\text{OUT},s} = \angle \left(\mathbf{w}_{\eta,L}^H \mathbf{R}_{\mathbf{xx}} \mathbf{w}_{\eta,R} \right) \, . \tag{38}$$

Since the speech ITF is unaffected by the filtering process, the binaural speech cues are also preserved independent of the value of η, i.e.,

$$\text{ILD}_{\text{OUT},s} = \text{ILD}_{\text{IN},s}, \quad \text{IPD}_{\text{OUT},s} = \text{IPD}_{\text{IN},s} \, . \tag{39}$$

The same analysis can be applied to the binaural noise cues where the noise ILD and IPD at the input are given by

$$\text{ILD}_{\text{IN},n} = |\text{ITF}_{\text{IN},n}| \tag{40}$$
$$= \frac{\mathbf{e}_L^T \mathbf{R}_{\mathbf{nn}} \mathbf{e}_L}{\mathbf{e}_R^T \mathbf{R}_{\mathbf{nn}} \mathbf{e}_R} \tag{41}$$
$$= \frac{P_{n,L}}{P_{n,R}} \tag{42}$$
$$\text{IPD}_{\text{IN},n} = \angle\left(\text{ITF}_{\text{IN},n}\right) \tag{43}$$
$$= \angle\left(\frac{\mathbf{e}_L^T \mathbf{R}_{\mathbf{nn}} \mathbf{e}_R}{\mathbf{e}_R^T \mathbf{R}_{\mathbf{nn}} \mathbf{e}_R}\right) \tag{44}$$
$$= \angle\left(\phi_n\right) \ . \tag{45}$$

However, as shown in (29), since the noise ITF is affected by the filtering process, the binaural noise cues will not be preserved. Using (12) the noise ILD and IPD at the output are given as

$$\text{ILD}_{\text{OUT},n} = \frac{\mathbf{w}_{\eta,L}^H \mathbf{R}_{\mathbf{nn}} \mathbf{w}_{\eta,L}}{\mathbf{w}_{\eta,R}^H \mathbf{R}_{\mathbf{nn}} \mathbf{w}_{\eta,R}} \tag{46}$$
$$= \frac{\frac{(1-\eta)^2 \lambda P_{s,L}}{(\mu+\lambda)^2} + \frac{2(1-\eta)\eta P_{s,L}}{(\mu+\lambda)} + \eta^2 P_{n,L}}{\frac{(1-\eta)^2 \lambda P_{s,R}}{(\mu+\lambda)^2} + \frac{2(1-\eta)\eta P_{s,R}}{(\mu+\lambda)} + \eta^2 P_{n,R}} \tag{47}$$
$$\text{IPD}_{\text{OUT},n} = \angle\left(\mathbf{w}_{\eta,L}^H \mathbf{R}_{\mathbf{nn}} \mathbf{w}_{\eta,R}\right) \tag{48}$$
$$= \angle\left(\frac{(1-\eta)^2 \lambda \phi_s}{(\mu+\lambda)^2} + \frac{2(1-\eta)\eta\phi_s}{(\mu+\lambda)} + \eta^2 \phi_n\right) . \tag{49}$$

Again, looking at the limiting cases of η results in

$$\text{ILD}_{\text{OUT},n} = \begin{cases} \text{ILD}_{\text{OUT},s} = \text{ILD}_{\text{IN},s} & : \eta = 0 \\ \text{ILD}_{\text{IN},n} & : \eta = 1 \end{cases} , \tag{50}$$

$$\text{IPD}_{\text{OUT},n} = \begin{cases} \text{IPD}_{\text{OUT},s} = \text{IPD}_{\text{IN},s} & : \eta = 0 \\ \text{IPD}_{\text{IN},n} & : \eta = 1 \end{cases} \tag{51}$$

which shows that the noise ILD and IPD at the output will be closer to the speech ILD and IPD at the input and output with lower values of η, whereas at higher values of η the noise ILD and IPD at the output are closer to the noise ILD and IPD at the input.

Note that while the input binaural noise cues are defined on the total noise component (27), (41) and (44), in the case of a single coherent noise source, we are truly only interested in preserving the binaural cues of this noise source. However, if it assumed that $\delta \ll P_v$ in (14), then the binaural noise cues are approximately equal to the spatially correlated noise source alone, i.e., the noise power can be approximated as $P_{n,\{L,R\}} \approx P_{v,\{L,R\}}$ and the phase component can be approximated as $\phi_n \approx \mathbf{e}_R^T \mathbf{v}\mathbf{v}^H \mathbf{e}_L$. In the case of a single coherent noise source, the formulas are therefore good approximations for the case when δ is small, as they lead to compact and interpretable derivations with results that also have perceptual relevance.

IV. Data model with a remote microphone signal

In the previous section it was shown that when using the MWF_η, the binaural noise cues will always be distorted for $\eta \neq 1$. In fact, as more emphasis is put on preserving the binaural noise cues, i.e., a higher value of η is used, this will result in a decrease in the output SNR.

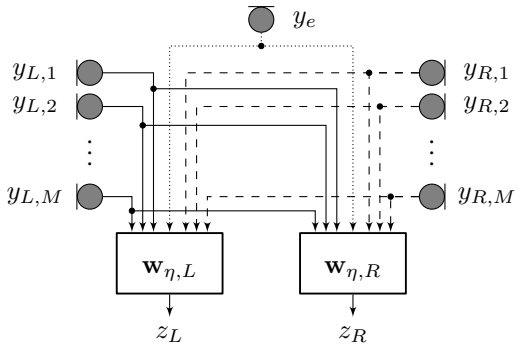


Fig. 2: The general binaural noise reduction system with a remote microphone signal.

We now suppose that a remote microphone signal, y_e, becomes available that can be used by both devices for the filtering process, where a depiction of this modified binaural noise reduction system is given in Figure 2. Unlike current binaural systems that mostly disable the local microphones when a remote microphone signal is available, we look to incorporate y_e into the noise reduction.

A. General data model with a remote microphone signal

We define the speech and noise steering vectors with this remote microphone signal included as $\mathbf{a}_e = \begin{bmatrix}\mathbf{a}^T \ a_e\end{bmatrix}^T$, $\mathbf{v}_e = \begin{bmatrix}\mathbf{v}^T \ v_e\end{bmatrix}^T$, respectively, and the new correlation matrices that incorporate these statistics as

$$\mathbf{R}_{\mathbf{x}_e\mathbf{x}_e} = \mathcal{E}\{\mathbf{x}_e\mathbf{x}_e^H\}, \ \mathbf{R}_{\mathbf{n}_e\mathbf{n}_e} = \mathcal{E}\{\mathbf{n}_e\mathbf{n}_e^H\} \tag{52}$$

where $\mathbf{x}_e = \begin{bmatrix}\mathbf{x}^T \ x_e\end{bmatrix}^T$ and $\mathbf{n}_e = \begin{bmatrix}\mathbf{n}^T \ n_e\end{bmatrix}^T$. The speech correlation matrix, which is still a rank-1 matrix, can be represented as

$$\mathbf{R}_{\mathbf{x}_e\mathbf{x}_e} = P_s \mathbf{a}_e \mathbf{a}_e^H \ . \tag{53}$$

Again, for ease of exposition we represent $\text{Tr}\{\mathbf{R}_{\mathbf{n}_e\mathbf{n}_e}^{-1}\mathbf{R}_{\mathbf{x}_e\mathbf{x}_e}\}$ as

$$\lambda_e \triangleq \text{Tr}\{\mathbf{R}_{\mathbf{n}_e\mathbf{n}_e}^{-1}\mathbf{R}_{\mathbf{x}_e\mathbf{x}_e}\} \tag{54}$$

where we show in the appendix, that with the addition of the remote microphone signal $\lambda_e \geq \lambda$.

B. Data model with single coherent noise source and a remote microphone signal

If a single coherent noise source is present, the noise correlation matrix can be represented similarly to (14) as

$$\mathbf{R}_{\mathbf{n}_e\mathbf{n}_e} = \mathbf{v}_e\mathbf{v}_e^H + \left[\begin{array}{c|c} \delta\mathbf{I} & \mathbf{0} \\ \hline \mathbf{0} & \delta_e \end{array}\right] \tag{55}$$

where δ_e is the incoherent noise power in the remote microphone signal. Using (53) and (55), λ_e can be also be given in a similar fashion as in (17) yielding

$$\lambda_e = P_s \left[\frac{||\mathbf{a}||^2}{\delta} + \frac{|a_e|^2}{\delta_e} - \frac{\left|\frac{\mathbf{a}^H\mathbf{v}}{\delta} + \frac{a_e^* v_e}{\delta_e}\right|^2}{1 + \frac{||\mathbf{v}||^2}{\delta} + \frac{|v_e|^2}{\delta_e}}\right] . \tag{56}$$

In the case that the incoherent noise power in the remote microphone signal is the same as the other microphones, e.g. if the remote and local AP microphones share similar characteristics, such that $\delta_e = \delta$, (56) reduces to

$$\lambda_e = \frac{P_s}{\delta}\left[||\mathbf{a}_e||^2 - \frac{|\mathbf{a}_e^H \mathbf{v}_e|^2}{\delta + ||\mathbf{v}_e||^2}\right] . \tag{57}$$

In the extreme case, when the remote microphone signal does not contain any noise component that is coherent with the noise in the other microphones, i.e., $v_e \to 0$ (56) can be given as

$$\lambda_e = \lambda + \frac{P_s|a_e|^2}{\delta_e}. \tag{58}$$

In the particular case when the speech and noise have similar steering vectors, an improvement in the output SNR is still possible, whereas the left and right array alone may not improve the output SNR as described in Section III-A . Note that, even if the remote microphone contains a noise component that is correlated to the coherent noise in the other microphones, the additional spatial diversity usually disambiguates the noise and the speech steering vector, i.e., $\frac{\mathbf{v}_e}{\|\mathbf{v}_e\|} \neq \frac{\mathbf{a}_e}{\|\mathbf{a}_e\|}$, even when $\frac{\mathbf{v}}{\|\mathbf{v}\|} \approx \frac{\mathbf{a}}{\|\mathbf{a}\|}$.

While extrapolating an exact increase in λ_e from (56) is intractable, some general observations can be made:

- There is an increase in λ_e with decreasing incoherent noise power on the remote microphone signal, δ_e.
- Conversely, increasing incoherent noise power on the remote microphone signal reduces the increase in λ_e.
- If the remote microphone signal decreases the collinearity between $\mathbf{a}$ and $\mathbf{v}$, i.e. the most right hand term in (57), this will result in an increase in λ_e.

Observing the extreme case, (58), the increase in λ_e comes from again either a decreasing incoherent noise power on the remote microphone signal, δ_e, or by increasing the magnitude squared of the remote microphone steering component, $|a_e|^2$, which could be accomplished by bringing the speech source and remote microphone closer together.

V. Performance measures with a remote microphone signal

In this section we show that the addition of a remote microphone signal, which results in $\lambda_e \geq \lambda$ as mentioned in Section IV, results in an increase in the output SNR up to a limiting point of $\frac{1}{\eta^2}$ and also increases the preservation of the binaural noise cues. Since the binaural speech cues obviously continue to be preserved for any value of η we will only analyze the performance of the output SNR and the binaural noise cues with a remote microphone signal.

A. Output SNR with a remote microphone signal

We first show that the output SNR with the addition of a remote microphone signal, $\text{SNR}^e_{\text{OUT},\{L,R\}}$, for a fixed value of μ and η is always greater than or equal to the output SNR without the remote microphone signal, i.e., $\text{SNR}^e_{\text{OUT},\{L,R\}} \geq \text{SNR}_{\text{OUT},\{L,R\}}$. This relies on the fact that $\lambda_e \geq \lambda$, as shown in the appendix.

The output SNR with the addition of a remote microphone signal is given the same as in (22) with λ_e substituted for λ, i.e.,

$$\text{SNR}^e_{\text{OUT},\{L,R\}} = \frac{\frac{(1-\eta)^2\lambda_e^2 P_{s,\{L,R\}}}{(\mu+\lambda_e)^2} + \frac{2(1-\eta)\eta\lambda_e P_{s,\{L,R\}}}{(\mu+\lambda_e)} + \eta^2 P_{s,\{L,R\}}}{\frac{(1-\eta)^2\lambda_e P_{s,\{L,R\}}}{(\mu+\lambda_e)^2} + \frac{2(1-\eta)\eta P_{s,\{L,R\}}}{(\mu+\lambda_e)} + \eta^2 P_{n,\{L,R\}}} . \tag{59}$$

The derivative of the numerator of (22) with respect to λ is given as

$$\frac{2\mu(1-\eta)(\eta\mu+\lambda)P_{s,\{L,R\}}}{(\mu+\lambda)^3} \geq 0 \tag{60}$$

which shows that the numerator is an increasing function with respect to λ. Therefore, the numerator of (59) will always be greater than or equal to the numerator of (22) since $\lambda_e \geq \lambda$. Likewise, by inspection, the denominator of (22) is always decreasing toward $\eta^2 P_{n,\{L,R\}}$ with increasing λ which means that the denominator of (59) will always be less than or equal to the denominator of (22) again since $\lambda_e \geq \lambda$. This then shows that $\text{SNR}^e_{\text{OUT},\{L,R\}} \geq \text{SNR}_{\text{OUT},\{L,R\}}$. However, if we look at the limiting case of $\lambda_e \to +\infty$ in (59) which approximately holds if $\lambda_e \gg \mu, \eta$ then we find

$$\lim_{\lambda_e \to +\infty} \text{SNR}^e_{\text{OUT},\{L,R\}} = \frac{\text{SNR}_{\text{IN},\{L,R\}}}{\eta^2} \tag{61}$$

which shows that with increasing values of λ_e, the closer the solution approaches the limiting SNR increase of $\frac{1}{\eta^2}$. This limiting case is in fact a byproduct of the original cost function (6), where the use of η essentially targets some residual noise containing the original binaural noise cues.

The output SNR_{OUT} can also be given in terms of SNR_{IN} as in [10] which is repeated here for convenience as

$$\text{SNR}^e_{\text{OUT},\{L,R\}} = \frac{(\eta\mu+\lambda_e)^2\text{SNR}_{\text{IN},\{L,R\}}}{(1-\eta)(2\eta\mu+\lambda_e+\eta\lambda_e)\text{SNR}_{\text{IN},\{L,R\}} + \eta^2(\mu+\lambda_e)^2} . \tag{62}$$

This can then be used to show the increase in SNR_{OUT} for SNR_{IN} and increasing λ_e. Figure 3 shows the SNR_{OUT} for -20 dB $\leq \text{SNR}_{\text{IN}} \leq$ 20 dB with various values of λ_e, where $\mu = 1$ and $\lambda_e \geq \text{SNR}_{\text{IN}}$. We see that for $\eta = 0$, $\text{SNR}_{\text{OUT}} = \lambda_e$ and for $\eta = 1$, $\text{SNR}_{\text{OUT}} = \text{SNR}_{\text{IN}}$ as given in (23). We also see that for $\eta \neq \{0, 1\}$, as λ_e increases it reaches the limiting point as given in (61).

B. Binaural noise cues with a remote microphone signal

A similar analysis can be performed with the binaural noise cues with a remote microphone signal. We now denote $\text{ITF}^e_{\text{OUT},n}$, $\text{ILD}^e_{\text{OUT},n}$ and $\text{IPD}^e_{\text{OUT},n}$ as the noise ITF, ILD and IPD at the output respectively with λ_e substituted for λ in (29), (47) and (49. In cases where $\lambda_e \geq \mu$, the ITF and ILD become hyperbolic functions[2] that asymptotically converge to the $\text{ITF}_{\text{IN},n}$ and $\text{ILD}_{\text{IN},n}$ respectively with increasing λ_e. This implies that with larger values of λ_e, the noise ITF will be better preserved and will approach $\frac{P_{n,L}}{\phi_n}$, i.e., since $\lambda_e \geq \lambda$

[2]Only values in the first quadrant are considered due to the positive values of the speech and noise powers.

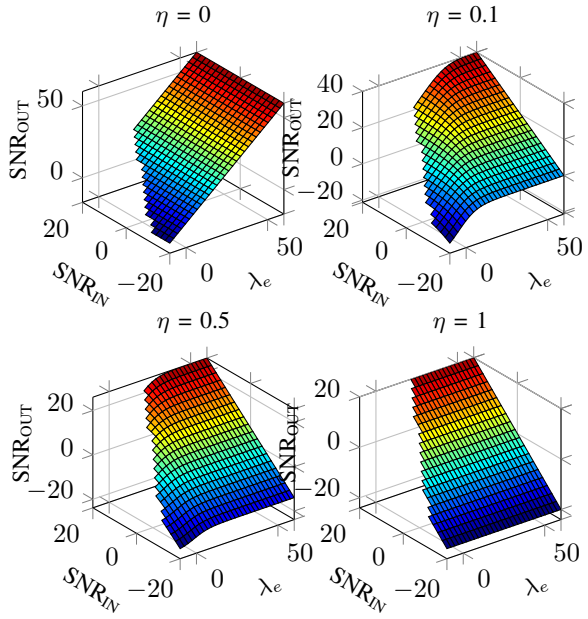


Fig. 3: SNR_{OUT} for different values of SNR_{IN} and λ_e.

the noise ITF will always be equivalent or better preserved with the addition of a remote microphone signal. This is also shown in Figure 4 for various values λ_e and μ where $\eta = 0.7$.

We can also look at a similar limiting case for the noise ITF with the addition of a remote microphone signal as $\lambda_e \to +\infty$. Using this assumption and substituting λ_e for λ in (29) yields

$$\lim_{\lambda_e \to +\infty} \text{ITF}^e_{\text{OUT},n} = \frac{P_{n,L}}{\phi_n} \tag{63}$$

$$= \text{ITF}_{\text{IN},n} \tag{64}$$

which shows that the noise ITF will be preserved.

The numerator and the denominator of the noise ILD (47), exhibit the same monotonic decrease with increasing λ_e, assuming that $\lambda_e >> \mu$ as the noise ITF. Therefore, it will also be better preserved for larger values of λ_e which, in the limiting case, results in

$$\lim_{\lambda_e \to +\infty} \text{ILD}^e_{\text{OUT},n} = \text{ILD}_{\text{IN},n} \ . \tag{65}$$

The first two terms of the numerator and denominator of the noise IPD in (49) both exhibit a monotonic decrease with increasing λ_e, which, in the limiting case, results in

$$\lim_{\lambda_e \to +\infty} \text{IPD}^e_{\text{OUT},n} = \text{IPD}_{\text{IN},n} \ . \tag{66}$$

It should be noted that in some extreme cases for small values of η and large values of μ such that $\mu > \lambda_e$, there may be a decrease in the preservation of the noise ITF and ILD before a similar asymptotic convergence is observed, as shown in Figure 5. However, as discussed in [25], μ is typically bounded by the desired signal distortion and input SNR of the signal and should be chosen as a function of the two, accordingly. Therefore, this decrease in the preservation of the

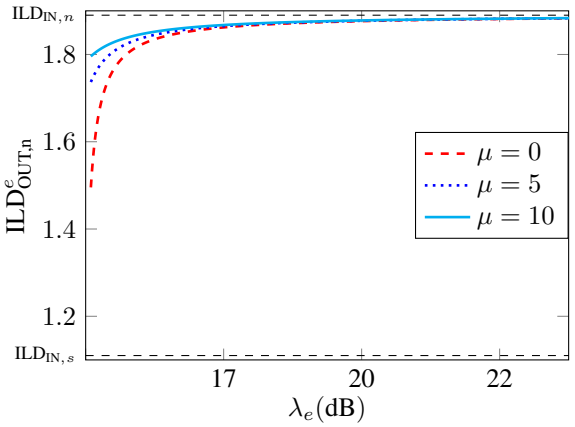


Fig. 4: The output noise ILD with a remote microphone signal for various values of λ_e and μ and where $\eta = 0.7$

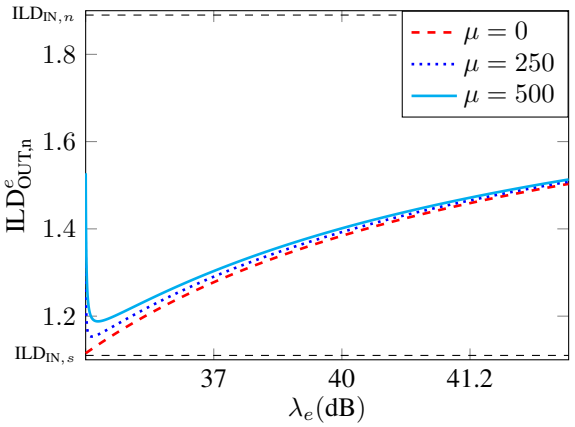


Fig. 5: The output noise ILD with a remote microphone signal for various values of λ_e and μ and where $\eta = 0.01$

noise ITF and ILD will not be observed as μ will be chosen in such a way that $\mu \leq \text{SNR}_{\text{IN}} \leq \lambda_e$.

If we define (12) using the correlation matrices containing the statistics with the remote microphone signal given in (52) then

$$\mathbf{w}_{\eta,L} = (1-\eta)\frac{\mathbf{R}^{-1}_{\mathbf{n}_e\mathbf{n}_e}\mathbf{R}_{\mathbf{x}_e\mathbf{x}_e}\mathbf{e}_L}{\mu + \text{Tr}\{\mathbf{R}^{-1}_{\mathbf{n}_e\mathbf{n}_e}\mathbf{R}_{\mathbf{x}_e\mathbf{x}_e}\}} + \eta\mathbf{e}_L \tag{67}$$

$$= (1-\eta)\mathbf{w}_{0,L} + \eta\mathbf{e}_L \tag{68}$$

where $\mathbf{w}_{0,L}$ is (67) with $\eta = 0$ (note $\mathbf{w}_{\eta,R}$ and $\mathbf{w}_{0,R}$ are defined in the same manner). We see that the first terms produces the speech with the correct binaural cues, plus residual noise with the wrong binaural cues, and the second term produces the correct binaural noise (and speech) cues, where the noise level is reduced by the target amount η. The function of the remote microphone signal is to improve the performance of $\mathbf{w}_{0,L}$ and $\mathbf{w}_{0,R}$, leading to less residual noise produced by the first term. As a consequence, $\mathbf{w}_{\eta,L}$ and $\mathbf{w}_{\eta,R}$ lead to better binaural noise cues due to the second term.

VI. Constrained PNE

In the previous section, it was shown that adding a remote microphone signal was able to increase the output, which in

the limiting case provided an SNR up to $\frac{1}{\eta^2}$, while simultaneously improving the binaural noise cues. In the derivations, it was assumed that η was constrained to the same value before and after the addition of the remote microphone signal. However, the PNE can be constrained is such a way such that η takes a new value of η_e and allows for further improvement of one performance measure while limiting the improvement of the other, e.g., some of the improvement in the binaural noise cues could be sacrificed to further increase the output SNR and vice versa.

Therefore, in this section, different methods of adjusting η to a new value of η_e are proposed while constraining the performance of either the output SNR or binaural noise cues. It is shown, that by constraining the binaural noise cues to the same performance before and after the addition of a remote microphone signal by adjusting the value of η to a new value of η_e, the output SNR is able to further be increased. Conversely, if the output SNR performance is constrained to the same value before and after the addition of a remote microphone signal by adjusting η to a new value of η_e, then the binaural noise cues are better preserved.

A. Increased output SNR

The noise ILD is constrained in order to adjust the value of η to a new value of η_e. The choice to constrain the noise ILD[3] is done for the ease of exposition, however, since $\text{ITF}^e_{\text{OUT},n}$, $\text{ILD}^e_{\text{OUT},n}$ and $\text{IPD}^e_{\text{OUT},n}$ all have the same dependence on η and λ_e, the change to η_e will affect all three identically. The new value of η_e is found by equating $\text{ILD}_{\text{OUT},n}(\eta)$ to $\text{ILD}^e_{\text{OUT},n}(\eta_e)$ which yields

$$\text{ILD}_{\text{OUT},n}(\eta) = \text{ILD}^e_{\text{OUT},n}(\eta_e) \tag{69}$$

$$= \frac{\frac{(1-\eta_e)^2\lambda_e P_{s,L}}{(\mu+\lambda_e)^2} + \frac{2(1-\eta_e)\eta_e P_{s,L}}{\mu+\lambda_e} + \eta_e^2 P_{n,L}}{\frac{(1-\eta_e)^2\lambda_e P_{s,R}}{(\mu+\lambda_e)^2} + \frac{2(1-\eta_e)\eta_e P_{s,R}}{\mu+\lambda_e} + \eta_e^2 P_{n,R}} \tag{70}$$

which is a quadratic expression in terms of η_e and has a closed-form solution that can easily be found with a standard symbolic solver. As shown in Section V, with the addition of a remote microphone signal, the binaural noise cues are better preserved, i.e., $\text{ILD}^e_{\text{OUT},n}$ will be closer to $\text{ILD}_{\text{IN},n}$ if the constraint of (69) is not present. This implies then, in order to reach the same performance before the addition of the remote microphone signal, $\text{ILD}_{\text{OUT},n}(\eta)$, that $\eta_e \leq \eta$. When plugging η_e in (59) this results in an increase in the output SNR compared to (59) with η. This method of adapting η, therefore, sacrifices the increase in the binaural noise cue preservation while simultaneously increasing the output SNR.

B. Increased binaural noise cue preservation

In a similar manner to Section VI-A, the value of η is adjusted to a new value of η_e based on constraining the output SNR. The new value of η_e is found by equating the output SNR without the remote microphone signal, $\text{SNR}_{\text{OUT},\{L,R\}}$, to the output SNR with the remote microphone signal, $\text{SNR}^e_{\text{OUT},\{L,R\}}$, yielding

$$\text{SNR}_{\text{OUT},\{L,R\}}(\eta) = \text{SNR}^e_{\text{OUT},\{L,R\}}(\eta_e) \tag{71}$$

$$= \frac{\frac{(1-\eta_e)^2\lambda_e^2 P_{s,\{L,R\}}}{(\mu+\lambda_e)^2} + \frac{2(1-\eta_e)\eta_e\lambda_e P_{s,\{L,R\}}}{(\mu+\lambda_e)} + \eta_e^2 P_{s,\{L,R\}}}{\frac{(1-\eta_e)^2\lambda_e P_{s,\{L,R\}}}{(\mu+\lambda_e)^2} + \frac{2(1-\eta_e)\eta_e P_{s,\{L,R\}}}{(\mu+\lambda_e)} + \eta_e^2 P_{n,\{L,R\}}} \tag{72}$$

which is again a quadratic expression in terms of η_e. Since, in Section V-A it was shown that $\text{SNR}^e_{\text{OUT},\{L,R\}}(\eta) \geq \text{SNR}_{\text{OUT},\{L,R\}}(\eta)$, then in order to achieve the same output SNR before and after the addition of the remote microphone signal $\eta_e \geq \eta$, thereby sacrificing the improvement in the output SNR. This larger value of η_e, however, simultaneously further improves the binaural noise cue preservation. Again since $\text{ITF}^e_{\text{OUT},n}(\eta_e)$, $\text{ILD}^e_{\text{OUT},n}(\eta_e)$ and $\text{IPD}^e_{\text{OUT},n}(\eta_e)$ all have the same dependence on η_e, this will lead to an equivalent preservation in all three binaural noise cues.

However, since $\text{SNR}_{\text{OUT},L} \neq \text{SNR}_{\text{OUT},R}$, $\forall \eta \neq 0$, the device with the smallest difference between $\text{SNR}_{\text{OUT},\{L,R\}}$ and $\text{SNR}^e_{\text{OUT},\{L,R\}}$ should be used as not to decrease the output SNR of either device to below the case without a remote microphone signal.

VII. Simulations

In this section we analyze the effect on the previously defined performance measures using the MWF_η with the addition of a remote microphone signal. The performance measures are also analyzed when η is adjusted based on constraining the PNE as described in Section VI. Simulations were performed using the head-related impulse responses (HRIRs) provided by the database in [30]. The majority of the simulations used the anechoic HRIRs provided, however, Section VII-B explores the case of a reverberant environment. The acoustic signals for the noise and speech were convolved with the HRIRs in order to produce the desired acoustic environment. The sampling frequency of the acoustic signals was 48 kHz, to match the sampling frequency of the HRIRs. Once the microphone signals had been generated, the signals were downsampled to 16 kHz. The microphone signals were split into windowed frames of length of $L = 512$ with a Hanning window and an overlap of $\frac{L}{2}$. The simulations were performed in batch mode, which means that the filters were found using the full length of the speech and noise signals to estimate $\mathbf{R}_{\mathbf{xx}}$, $\mathbf{R}_{\mathbf{x}_e\mathbf{x}_e}$, $\mathbf{R}_{\mathbf{nn}}$ and $\mathbf{R}_{\mathbf{n}_e\mathbf{n}_e}$. In real-time scenarios, however, the devices typically use a VAD as described in Section II, to gather the necessary statistics. To ensure the rank-1 assumption was held for both the $\mathbf{R}_{\mathbf{xx}}$ and $\mathbf{R}_{\mathbf{x}_e\mathbf{x}_e}$ matrices, the low rank approximation presented in [27] based on the generalized eigenvalues of the speech-and-noise and the noise correlation matrix was used.

The location of the speech and noise source are represented as θ_s and θ_n respectively, where $\theta_{\{s,n\}} = 0^\circ$ indicates the source was directly in front of the devices and $\theta_{\{s,n\}}$ increases in a clock-wise direction, e.g., $\theta_{\{s,n\}} = 90^\circ$ indicates the source was at the right ear. The speech and noise were positioned at a distance of 3 m and the microphone setup consisted of three behind-the-ear (BTE) microphones on each device,

[3] Note that instead of equating $\text{ILD}_{\text{OUT},n}(\eta)$ and $\text{ILD}^e_{\text{OUT},n}(\eta_e)$ in order to find η_e, a predefined threshold for the noise ILD could also be used to ensure a certain level of binaural noise cue preservation.

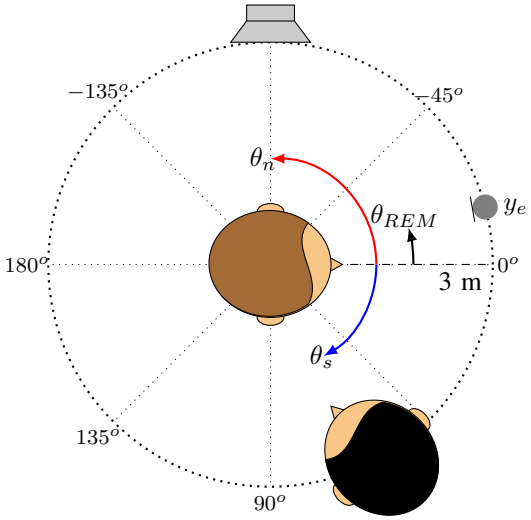


Fig. 6: A depiction of the speech source, noise source, and remote microphone, y_e, in terms of θ_s, θ_n and θ_{REM}.

$M = 3$, located at the front, middle and rear. Independent identically distributed noise that was equal to -10 dB of the coherent noise source signal power was also added to the signals. The trade-off parameter μ was set to the same value for all simulations, namely $\mu = 1$.

The remote microphone was assumed to be at the same distance from the listener as the speech and noise source where its location is given as θ_{REM}. A depiction of the speech source, noise source, and remote microphone in terms of θ_s, θ_n and θ_{REM} is given in Figure 6.

Using the location based on θ_{REM}, the distance between the speech and noise source was calculated and the remote microphone signal was generated using the image method proposed in [31]. The remote microphone was placed at different θ_{REM} resulting in different distances to the speech and noise and produced various input SNRs denoted as $\text{SNR}_{\text{IN},REM}$. In the case of the reverberant environment, presented in Section VII-B, the room was modeled with the same dimensions and T_{60}, namely 300 ms, as described in [30]. The additive noise of the remote microphone signal was equal to that of the other microphones in the system such that $\delta_e = \delta$. All narrowband results are shown for a frequency of $\omega = 1500$ Hz. This frequency is roughly the cut-off when IPDs become less important and ILDs become more important this frequency gives a good overview of the performance with respect to both binaural noise cues.

The performance measures presented in the previous sections are used in order to find the improvement in the output SNR and the binaural noise cues. The binaural speech cues are not analyzed as they are perfectly preserved as shown with the equivalence of the speech ITF (26). We define the improvement in the SNR for the left and right devices as

$$\Delta\text{SNR}_{\{L,R\}} = \text{SNR}^{e}_{\text{OUT},\{L,R\}} - \text{SNR}_{\text{IN},\{L,R\}} \ . \tag{73}$$

The noise ILD error is given as

$$\Delta\text{ILD}_n = 10\log_{10}(\text{ILD}^{e}_{\text{OUT},n}) - 10\log_{10}(\text{ILD}_{\text{IN},n}) \tag{74}$$

and a relative noise IPD error is given as

$$\Delta\text{IPD}_n = \frac{|\text{IPD}^{e}_{\text{OUT},n} - \text{IPD}_{\text{IN},n}|}{\pi} \tag{75}$$

where $0 \leq \Delta\text{IPD}_n \leq 1$.

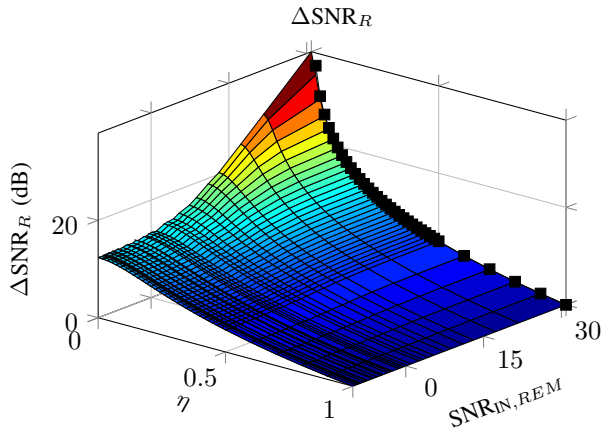
A. Performance measures with changing η and $\text{SNR}_{IN,REM}$

The first simulation was performed in order to highlight the benefit of using a remote microphone signal with a high $\text{SNR}_{\text{IN},REM}$. The speech and noise sources were located at $\theta_s = 30°$ and $\theta_n = -90°$ where the input SNR of the reference microphone of the right device was $\text{SNR}_{\text{IN},R} = 0$ dB. The position of the remote microphone signal was swept between $-10° \leq \theta_{REM} \leq 29°$ resulting in an input SNR at the remote microphone signal of approximately -2 dB $\leq \text{SNR}_{\text{IN},REM} \leq 40$ dB. The value for η was varied from 0 (no PNE) to 1 (no noise reduction). Figure 7 shows the narrowband ΔSNR_R, the noise ILD error and the relative noise IPD error.

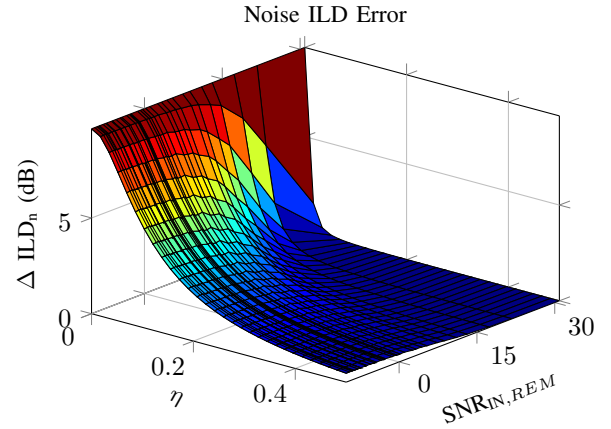
We see that for small values of η the remote microphone signal is able to provide a large increase in ΔSNR_R, while at larger values of η, since almost no noise reduction is performed, there is little to no improvement of ΔSNR_R. We see also that ΔSNR_R reaches the limiting case of $\frac{1}{\eta^2}$ (represented by ■ in Figure 9a) for large values of η even with a low input SNR of the remote microphone signal. However, at lower values of η, the remote microphone signal is able to increase the output SNR performance of the system to the limiting case of $\frac{1}{\eta^2}$. We see that for small values of η the largest error is produced as there is very little PNE. The errors become smaller with increasing η and $\text{SNR}_{\text{IN},REM}$. Unlike Figure 9a, η is in the range of $0 \leq \eta \leq 0.5$ as larger values offer little benefit to those already shown. We note that while the broadband $\text{SNR}_{\text{IN},REM}$ was in the range of $\approx$ -2 dB and 40 dB, the range of the narrowband $\text{SNR}_{\text{IN},REM}(\omega)$ is somewhat lower. Figure 8 show ΔSNR_R, the noise ILD error and the relative noise IPD error for all frequencies up to the Nyquist rate of 8000 Hz without the remote microphone signal (a,c,e) and with the remote signal at approximately 40 dB. We see that besides providing a large increase in ΔSNR_R for all frequencies, the noise ILD error and the relative noise IPD error both decrease at a substantial rate for even low values of η.

B. Performance measures with changing η and $\text{SNR}_{IN,REM}$ in a reverberant environment

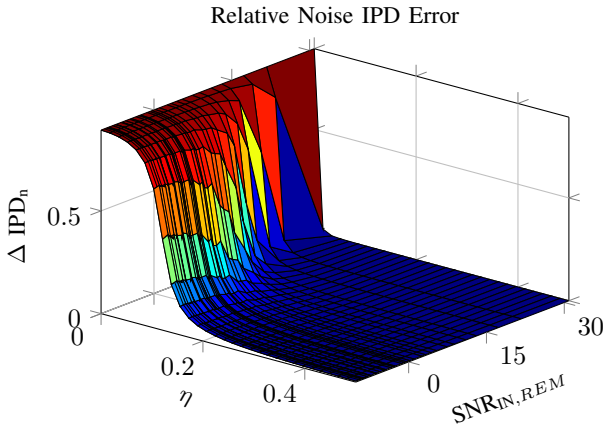
A reverberant environment is also available in the database provided in [30]. The speech and noise were positioned at a distance of 1 m and the same $\theta_{\{s,n\}}$ as the anechoic environment, namely $\theta_s = 30°, \theta_n = -90°$. Again the remote microphone was generated in the same manner as in Section VII where a $\text{T}_{60} = 300$ ms was used and the room dimensions were as close to the database environment as possible. The position of the remote microphone signal was swept in the same manner as in the previous section resulting in an input SNR at the remote microphone signal of approximately -4 dB $\leq \text{SNR}_{\text{IN},REM} \leq 25$ dB. This decrease in SNR was most likely a result of the speech and noise source



(a) ΔSNR_R for various values of η and $\text{SNR}_{\text{IN,REM}}$ along with the limiting case of $\frac{1}{\eta^2}$.



(b) Noise ILD error for various values of η and $\text{SNR}_{\text{IN},REM}$.



(c) Relative noise IPD error for various values of η and $\text{SNR}_{\text{IN},REM}$.

Fig. 7: ΔSNR_R with the limiting case of $\frac{1}{\eta^2}$ ■ (a), the noise ILD error (b) and the relative noise IPD error (c) for various values of η and $\text{SNR}_{\text{IN,REM}}$.

being closer together (1 m away from listener instead of 3 m) and not a product of the reverberant environment.

Figure 9 shows ΔSNR_R with the limiting case of $\frac{1}{\eta^2}$ ■, the noise ILD error and the relative noise IPD error for various values of η and $\text{SNR}_{\text{IN,REM}}$ in the reverberant environment. We see that the behavior essentially mimics that of the anechoic case where higher $\text{SNR}_{\text{IN,REM}}$ provides better output SNR as well as better preservation of the binaural noise cues for $\eta \neq 0$.

C. η based on constrained PNE

We now perform simulations where η is adjusted to a new value η_e, as described in Section VI based on constrained PNE where a numerical solver was used to find the solutions to (69) and (71).

We first find η_e by constraining the ILD as described in Section VI-A which allows for an increase in the output SNR. Figure 10 shows the adapted η_e, the $\Delta\text{SNR}_{\text{OUT},R}$, the noise ILD errors and the relative noise IPD errors with and without the adapted η for various values of η and $\text{SNR}_{\text{IN},REM}$. We see that by sacrificing the improvement in the binaural cue preservation, ΔSNR_R is further improved (the same holds true for ΔSNR_L). However, as shown in Figures 10c and 10d, this comes at the expense of keeping the binaural noise cues the same with and without the remote microphone signal.

We next find η_e by constraining the SNR as described in Section VI-B which allows for an increase in the binaural noise cue preservation. The output SNR of the left device was used as the constraint as $\text{SNR}^e_{\text{OUT},L} - \text{SNR}_{\text{OUT},L} < \text{SNR}^e_{\text{OUT},R} - \text{SNR}_{\text{OUT},R}$ which guarantees that the output SNR of either device is not reduced to below $\text{SNR}_{\text{OUT},\{L,R\}}$. Figure 11 shows the adapted η_e, the $\Delta\text{SNR}_{\text{OUT},R}$, the difference in the output noise ILD errors, $\Delta\text{ILD}_{\text{OUT},n} = \text{ILD}^e_{\text{OUT},n} - \text{ILD}_{\text{OUT},n}$ with the adapted η for various values of η and $\text{SNR}_{\text{IN},REM}$. We see that for most values of η and input SNR at the remote microphone signal there is only a marginal adaption. This lack of adaptation is typically the result of a large increase in the output SNR of the device that has the lowest input SNR, due to using signals from the higher input SNR device, and results in the device reaching the $\frac{1}{\eta^2}$ even without a remote microphone signal. However, for low values of η and before reaching the $\frac{1}{\eta^2}$, there are improvements in the output noise ILD error.

D. Influence of noise source position

Another interesting aspect is to observe the effect on the binaural noise cues for different θ_n values relative to θ_s. The speech source was placed at $\theta_s = 30°$ at .8 m and the noise source position was swept in the range of $-90° \leq \theta_n < 90°$ in increments of $5°$ at 3 m. The ΔSNR_R without a remote microphone signal, with a remote microphone signal where $\eta = 0.2$, which was shown to provide good localization for both the binaural speech and noise cues in [19] and finally where η was adapted based on constrained PNE using the ILD cues. The remote microphone signal was placed at $\theta_{REM} = 29°$ and at the same distance as the speech source.

Figure 12 shows ΔSNR_R versus various values of θ_n. We see that without the remote microphone signal, there is a mean

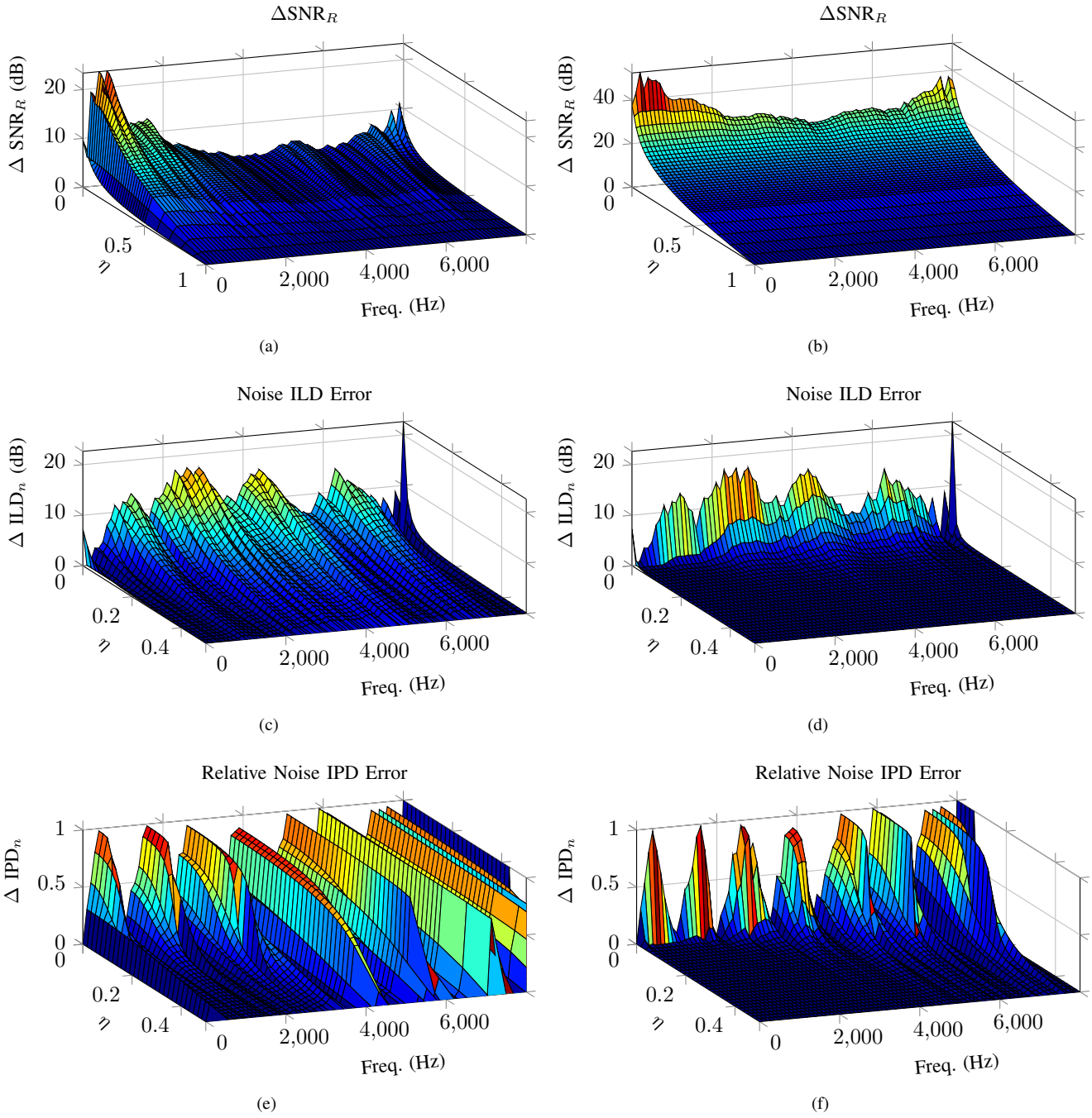
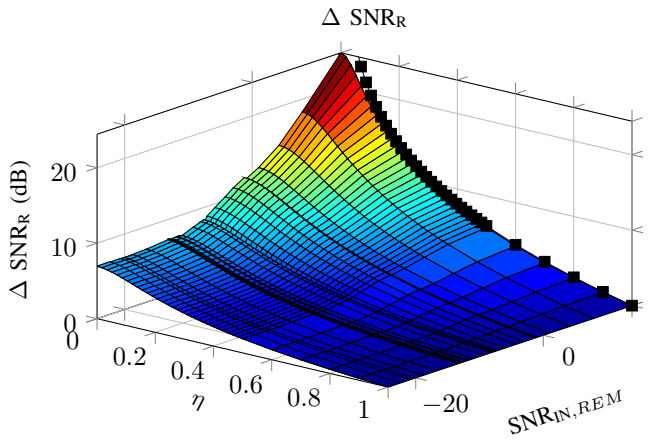


Fig. 8: ΔSNR_R, the noise ILD error and relative noise IPD for various frequencies and values of η without a remote microphone signal (a,c,e) and with a remote microphone signal (b,d,f) with $\text{SNR}_{\text{IN},REM} \approx 40$ dB.

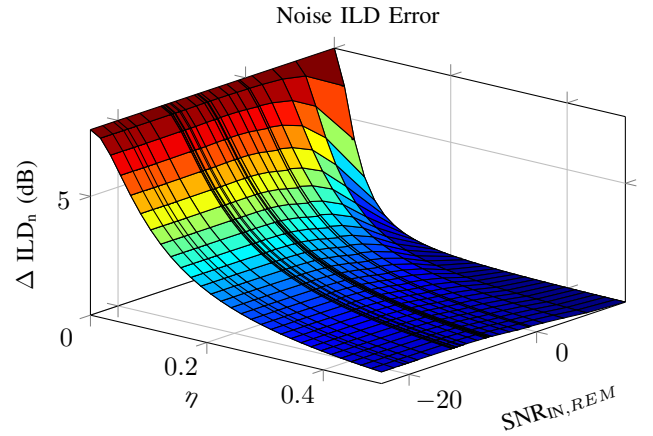
improvement of 2.7 dB, that degrades the closer θ_n gets to θ_s. We see that with the addition of the remote microphone signal there is a significant SNR increase, even when $\theta_s \approx \theta_n$. In using η_e, where the improvement of the binaural noise cues is sacrificed, we see a further improvement in ΔSNR_R. Again this comes at the expense of constraining the ILD to the same performance before and after the addition of the remote microphone signal.

VIII. Conclusions

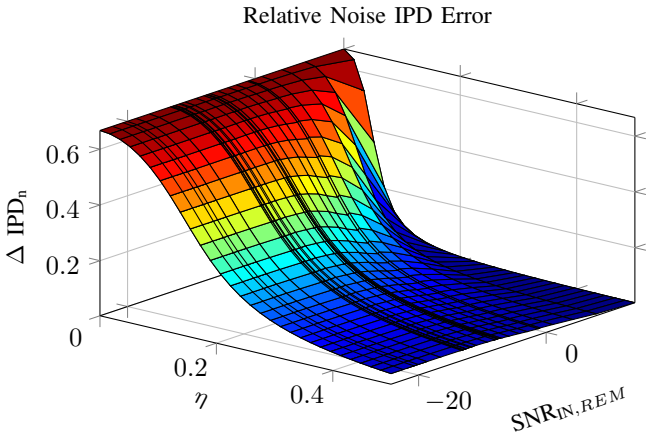
A general binaural noise reduction system was presented that used the MWF$_\eta$ in order to preserve the binaural speech and noise cues with the addition of a remote microphone signal. We showed that the addition of this remote microphone signal not only resulted in an increase in the output SNR, but also simultaneously improved the binaural noise cue preservation. A modification was proposed allowing the PNE variable η to change to a new value of η_e by exploiting the trade-off between noise reduction performance and binaural



(a) ΔSNR_R for various values of η and $\text{SNR}_{\text{IN,REM}}$ along with the limiting case of $\frac{1}{\eta^2}$.



(b) Noise ILD error for various values of η and $\text{SNR}_{\text{IN},REM}$.



(c) Relative noise IPD error for various values of η and $\text{SNR}_{\text{IN},REM}$.

Fig. 9: ΔSNR_R with the limiting case of $\frac{1}{\eta^2}$ ■ (a), the noise ILD error (b) and the relative noise IPD error (c) for various values of η and $\text{SNR}_{\text{IN,REM}}$ in a reverberant environment.

noise cue preservation. This allowed for an increase in one of the performance measures while constraining the other to the same value before and after the addition of the remote microphone signal.

APPENDIX A

In this appendix we show that adding a remote microphone signal results in $\lambda_e \geq \lambda$. We note that while in [29] it was shown that when there is no coherent noise, i.e., $\mathbf{v} = 0$, $\lambda_e \geq \lambda$, this was not shown for the general case with the addition of a coherent noise source.

We define a matrix pencil as

$$\mathbf{R}_{\mathbf{x}_e\mathbf{x}_e} - \lambda_e \mathbf{R}_{\mathbf{n}_e\mathbf{n}_e} \tag{76}$$

where $\mathbf{R}_{\mathbf{x}_e\mathbf{x}_e}$ is a Hermitian positive semi-definite matrix and $\mathbf{R}_{\mathbf{n}_e\mathbf{n}_e}$ is a Hermitian positive definite matrix. As $\mathbf{R}_{\mathbf{x}_e\mathbf{x}_e}$ is a rank-1 matrix, there will be only one nonzero eigenvalue, λ_e which is then in the same form as (57), repeated here for convenience

$$\text{Tr}\{\mathbf{R}_{\mathbf{n}_e\mathbf{n}_e}^{-1}\mathbf{R}_{\mathbf{x}_e\mathbf{x}_e}\} = \lambda_e \ . \tag{77}$$

Let $\mathbf{C}$ be a $(2M+1) \times 2M$-dimensional matrix defined as

$$\mathbf{C} = \begin{bmatrix} \mathbf{I} \\ \mathbf{0}^T \end{bmatrix} \tag{78}$$

where $\mathbf{I}$ is an $2M \times 2M$-dimensional identity matrix and $\mathbf{0}$ is a $2M$-dimensional row vector of all zeros. Using (78) we define the so-called *Rayleigh Quotient Pencil* as

$$\mathbf{C}^T\mathbf{R}_{\mathbf{x}_e\mathbf{x}_e}\mathbf{C} - \lambda\mathbf{C}^T\mathbf{R}_{\mathbf{n}_e\mathbf{n}_e}\mathbf{C} \tag{79}$$

$$= \mathbf{R}_{\mathbf{xx}} - \lambda\mathbf{R}_{\mathbf{nn}} \ . \tag{80}$$

As $\mathbf{R}_{\mathbf{xx}}$ is again a rank-1 matrix, it will have only one nonzero eigenvalue λ, which can be represented as

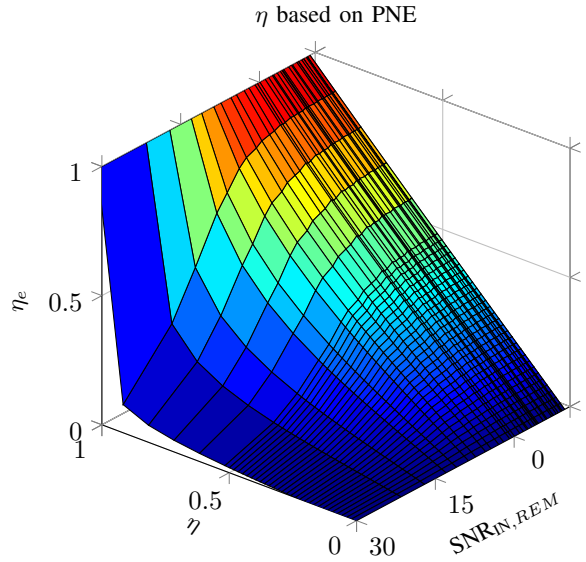
$$\text{Tr}\{\mathbf{R}_{\mathbf{nn}}^{-1}\mathbf{R}_{\mathbf{xx}}\} = \lambda \ . \tag{81}$$

Since $\mathbf{C} \in \mathbb{R}^{(2M+1)\times 2M}$ with $\text{rank}(\mathbf{C}) = 2M$ then, as shown in [32], by way of Cauchy interlacing inequalities the eigenvalue of the Rayleigh Quotient Pencil follows

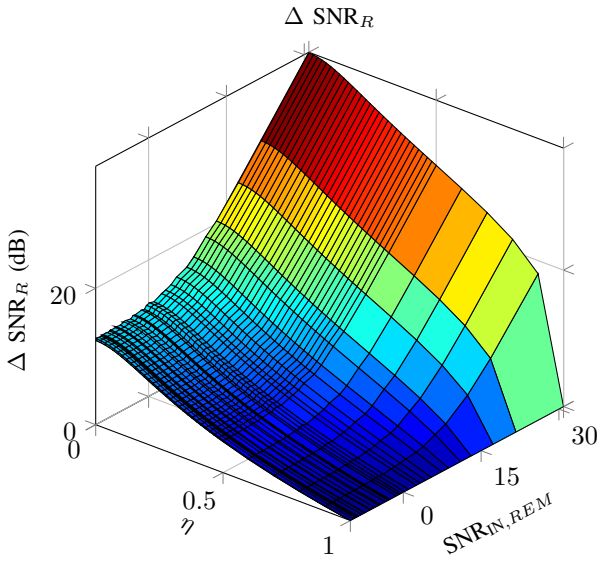
$$\lambda \leq \lambda_e \ . \tag{82}$$

REFERENCES

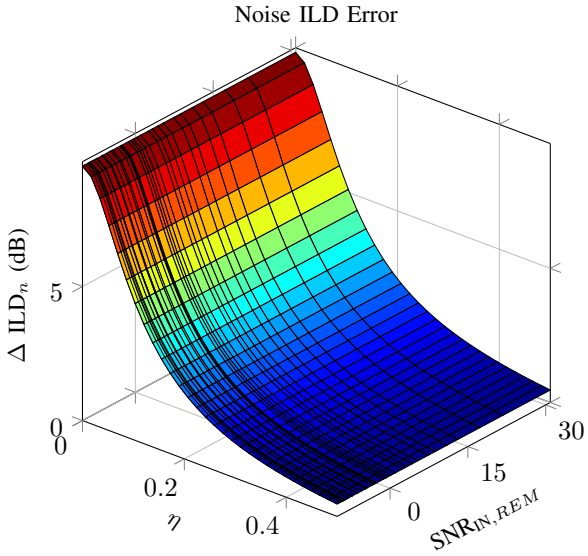
[1] M. C. Killion, "SNR loss : I can hear what people say, but I cant understand them," *Hearing Review*, vol. 12, pp. 8–14, Apr. 1997.
[2] H. Dillon, *Hearing Aids*. Thieme, 2012.
[3] J. Wouters and J. Vanden Berghe, "Speech recognition in noise for cochlear implantees with a two-microphone monaural adaptive noise reduction system," *Ear and Hearing*, vol. 22, no. 5, pp. 420–430, Oct. 2001.
[4] J. G. Desloge, W. M. Rabinowitz, and P. M. Zurek, "Microphone-array hearing aids with binaural output - Part I : Fixed-processing systems," *IEEE Trans. on Audio, Speech, and Language Process.*, vol. 5, no. 6, pp. 529–542, Nov. 1997.
[5] J. Wouters, S. Doclo, R. Koning, and T. Francart, "Sound processing for better coding of monaural and binaural cues in auditory prostheses," *Proc. of the IEEE*, vol. 101, no. 9, pp. 1986–1997, Sep. 2013.
[6] D. Wang and G. J. Brown, *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*. Wiley-IEEE Press, 2006.
[7] R. Gilkey and T. R. Anderson, *Binaural and Spatial Hearing in Real and Virtual Environments*. Lawrence Erlbaum, 1997.



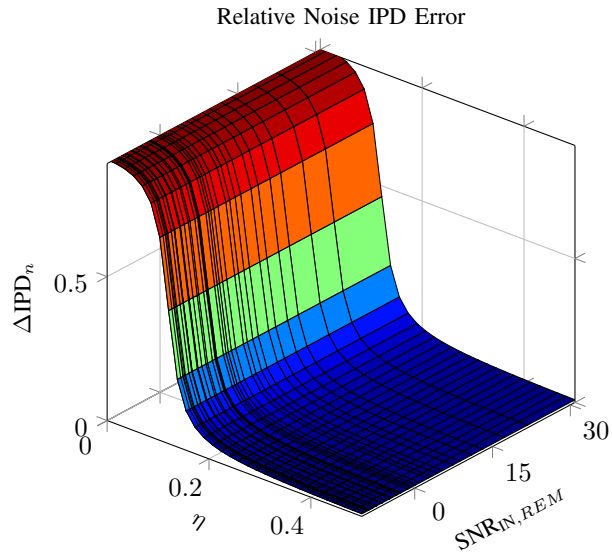
(a) Changing η based on ILD-constrained PNE.



(b) ΔSNR_R with changing η where the target $\frac{1}{\eta^2}$ is exceeded.



(c) Noise ILD error with changing η.



(d) Relative noise IPD with changing η.

Fig. 10: Changing η based on ILD-constrained PNE comparing ΔSNR_R, the noise ILD and relative noise IPD errors for various values of η_e and $\text{SNR}_{\text{IN},REM}$ (Note that the x-y axis of (a) are flipped for an easier graphical representation).

[8] S. Doclo, T. J. Klasen, T. Van den Bogaert, J. Wouters, , and M. Moonen, "Theoretical analysis of binaural cue preservation using multi-channel Wiener filtering and interaural transfer functions," in *Proc. of the Int. Workshop on Acoustic Signal Enhancement (IWAENC)*, Paris, France, Sep. 2006.

[9] T. Van den Bogaert, J. Wouters, S. Doclo, and M. Moonen, "Binaural cue preservation for hearing aids using an interaural transfer function multichannel Wiener filter," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, vol. 4, Apr. 2007, pp. 565–568.

[10] B. Cornelis, S. Doclo, T. Van den Bogaert, M. Moonen, and J. Wouters, "Theoretical analysis of binaural multimicrophone noise reduction techniques," *IEEE Trans. on Audio, Speech, and Language Process.*, vol. 18, no. 2, pp. 342–355, Feb 2010.

[11] R. H. Gifford and L. J. Revit, "Speech perception for adult cochlear implant recipients in a realistic background noise: Effectiveness of preprocessing strategies and external options for improving speech recognition in noise," *Journal of the American Academy of Audiology*, vol. 21.7, pp. 441–488, Jul.-Aug. 2010.

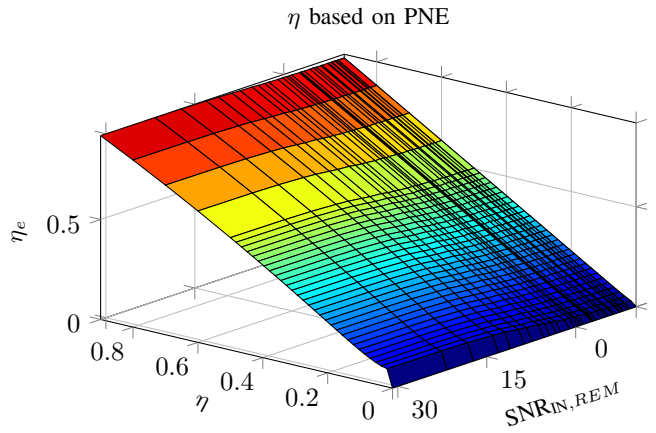
[12] E. C. Schafer and L. M. Thibodeau, "Electroacoustic verification of FM benefits in advanced hearing aid circuitry," *The Journal of the Acoustical Society of America*, vol. 112, no. 5, pp. 2271–2272, 2002.

[13] A. Bertrand and M. Moonen, "Robust distributed noise reduction in hearing aids with external acoustic sensor nodes," *EURASIP J. Adv. Sig. Process.*, 2009.

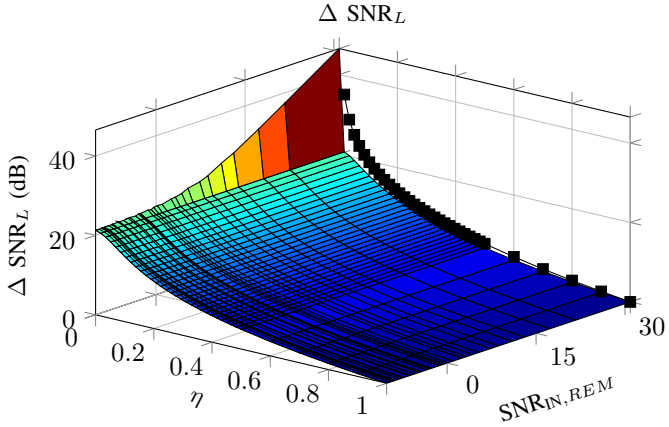
[14] E. C. Schafer, K. Sanders, D. Bryant, K. Keeney, and N. Baldus, "Effects of voice priority in FM systems for children with hearing aids," *Journal of Educational Audiology*, vol. 19, pp. 12–24, 2013.

[15] G. Grimm, G. Guilmin, F. Poppen, M. Vlaming, and V. Hohmann, "The personal hearing system–a software hearing aid for a personal communication system," *EURASIP J. Adv. Sig. Process.*, no. 1, pp. 1–9, 2009.

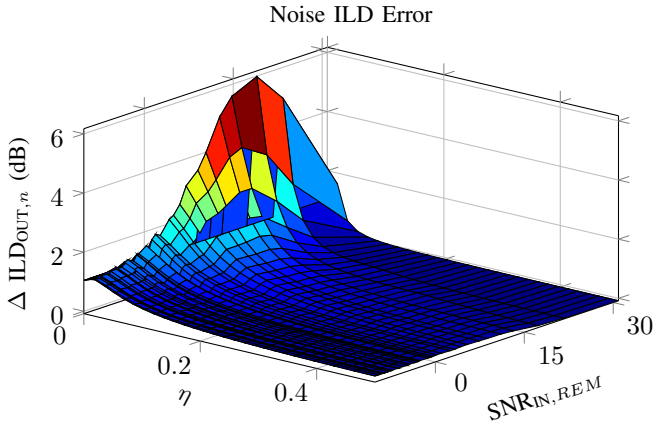
[16] J. Wolfe, M. Morais, E. Schafer, E. Mills, H. E. Mlder, F. Gold-



(a) Changing η based on constraining ΔSNR_L.

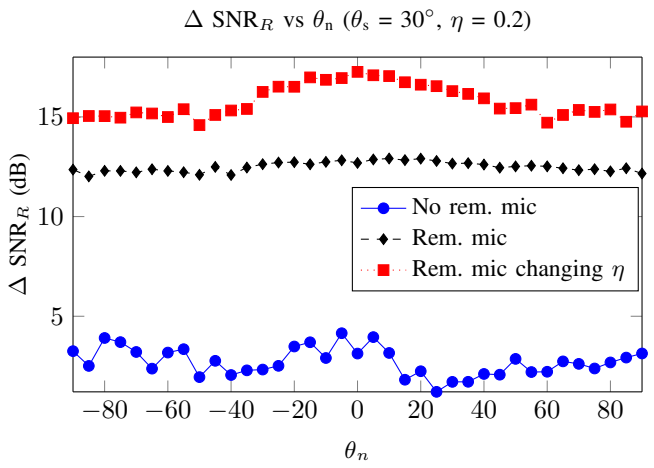


(b) ΔSNR_L with constraining η_e.



(c) ΔSNR_L with constraining η_e.

Fig. 11: Changing η based on the ΔSNR_L PNE constraint and the resulting $\Delta\text{SNR}_L^e(\eta_e)$ (Note that the x-y axis of (a) are flipped for an easier graphical representation).



(a) ΔSNR_R for various values of θ_n.

Fig. 12: The ΔSNR_R for a single frequency ($\omega = 1500$ Hz) for various values of θ_n where $\theta_s = 30°$ and η is at a constant value, namely, $\eta = 0.2$ for the scenarios with no remote microphone and with a remote microphone with fixed η.

beck, F. Marquis, A. John, M. Hudson, B. R. Peters, and L. Lianos, "Evaluation of speech recognition of cochlear implant recipients using a personal digital adaptive radio frequency system," *Journal of the American Academy of Audiology*, vol. 24, no. 8, pp. 714–724, 2013.

[17] T. Francart, A. Lenssen, and J. Wouters, "Enhancement of interaural level differences improves sound localization in bimodal hearing," *The Journal of the Acoustical Society of America*, vol. 130, no. 5, pp. 2817–2826, 2011.

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

[19] T. Van den Bogaert, S. Doclo, J. Wouters, and M. Moonen, "The effect of multimicrophone noise reduction systems on sound source localization by users of binaural hearing aids," *The Journal of the Acoustical Society of America*, vol. 124, no. 1, pp. 484–497, 2008.

[20] ——, "Speech enhancement with multichannel Wiener filter techniques in multimicrophone binaural hearing aids," *The Journal of the Acoustical Society of America*, vol. 125, no. 1, pp. 360–371, 2009.

[21] T. C. Lawin-Ore and S. Doclo, "Reference microphone selection for MWF-based noise reduction using distributed microphone arrays," in *Proc. 10th ITG Symp. on Speech Commun.*, Sep. 2012, pp. 1–4.

[22] S. Stenzel, T. C. Lawin-Ore, J. Freudenberger, and S. Doclo, "A multichannel Wiener filter with partial equalization for distributed microphones," in *Proc. IEEE Workshop Appl. of Signal Process. to Audio and Acoustics (WASPAA)*, Oct. 2013, pp. 1–4.

[23] Y. Hu and P. C. Loizou, "A generalized subspace approach for enhancing speech corrupted by colored noise," *IEEE Trans. on Speech and Audio Process.*, vol. 11, no. 4, pp. 334–341, Jul. 2003.

[24] B. Defraene, K. Ngo, T. van Waterschoot, M. Diehl, and M. Moonen, "A psychoacoustically motivated speech distortion weighted multi-channel Wiener filter for noise reduction," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, Mar. 2012, pp. 4637–4640.

[25] S. Braun, K. Kowalczyk, and E. A. P. Habets, "Residual noise control using a parametric multichannel Wiener filter," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, Apr. 2015, pp. 360–364.

[26] D. Marquardt, V. Hohmann, and S. Doclo, "Perceptually motivated coherence preservation in multi-channel Wiener filtering based noise reduction for binaural hearing aids," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, May 2014, pp. 3660–3664.

[27] R. Serizel, M. Moonen, B. Van Dijk, and J. Wouters, "Low-rank approximation based multichannel Wiener filter algorithms for noise reduction with application in cochlear implants," *IEEE/ACM Trans. on Audio, Speech, and Language Process.*, vol. 22, no. 4, pp. 785–799, Apr. 2014.

[28] J. Benesty, S. Makino, and J. Chen, *Speech Enhancement*, ser. Signals and Communication Technology. Springer Berlin Heidelberg, 2005.
[29] M. Souden, J. Benesty, and S. Affes, "On optimal frequency-domain multichannel linear filtering for noise reduction," *IEEE Trans. on Audio, Speech, and Language Process.*, vol. 18, no. 2, pp. 260–276, Feb. 2010.
[30] H. Kayser, S. D. Ewert, J. Anemüller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses," *EURASIP J. Adv. Sig. Process.*, vol. 2009, pp. 1–10, 2009.
[31] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, pp. 943–950, 1979.
[32] R.-C. Li, "On eigenvalue variations of Rayleigh quotient matrix pencils of a definite pencil," *Linear Algebra and its Applications*, vol. 208209, pp. 471 – 483, 1994.



Joseph Szurley received his B.A. degree in physics from the University of California at Berkeley (2005) his M.Sc. in electrical engineering with honors from California State University Long Beach (2010) and his PhD from KU Leuven, Belgium (2015). His research interests include audio signal processing, wireless (acoustic) sensor networks and distributed signal and speech enhancement. In 2003 he worked at Lawrence Berkeley National Laboratory in the Atomic Molecular and Optical Sciences (AMSO) division. From 2004-2007 he was a research assistant for Takeda Pharmaceuticals. From 2009-2010 he worked at Conexant Systems, Inc. in the product engineering group. Currently he is a postdoctoral researcher at KU Leuven focusing on noise reduction strategies for cochlear implants.



Alexander Bertrand (M'08) received the Ph.D. degree in Engineering Sciences (Electrical Engineering) from KU Leuven, Belgium, in 2011. He is currently assistant professor with the Electrical Engineering Department, KU Leuven. He was a visiting researcher at University of California, Los Angeles, at University of California, Berkeley, and at Imec-NL Eindhoven (The Netherlands). Dr. Bertrand received the 2012 FWO/IBM-Belgium Award, the 2013 KU Leuven Research Council Award, and an Honorary Mention in the 2013 ERCIM Cor Baayen Award. He has served as a technical program committee member for several signal processing conferences and workshops (EUSIPCO, IEEE GlobalSIP, IEEE WASPAA), and as lead guest editor for Signal Processing (Elsevier). His main research focus is on biomedical signal processing, sensor arrays and distributed signal processing, with applications in EEG, hearing prostheses, neuroprobes, and microphone arrays.



Bas van Dijk got his PhD in biophysics at the University of Leiden (the Netherlands) in 1997. After that he started working in the cochlear implant clinic of the University Medical Hospital in Utrecht (the Netherlands) mainly on research projects on objective fitting of cochlear implants as well as doing clinical audiological work (selection and fitting) with cochlear implant recipients. In 2001 he joined Cochlear, initially as a clinical advisor to the clinical software team, later as research coordinator. Today Bas is responsible for Cochlears sound coding/signal processing research projects worldwide. He worked on many innovations in the Cochlear system (Beam, AutoNRT, Nucleus Fitting Software, Remote Assistant Fitting, SNR-NR,etc) , is the holder of a dozen patents and (co-)author of about fifteen scientific papers in the field of cochlear implants.



Marc Moonen (M'94, SM'06, F'07) is a Full Professor at the Electrical Engineering Department of KU Leuven, where he is heading a research team working in the area of numerical algorithms and signal processing for digital communications, wireless communications, DSL and audio signal processing.
He received the 1994 KU Leuven Research Council Award, the 1997 Alcatel Bell (Belgium) Award (with Piet Vandaele), the 2004 Alcatel Bell (Belgium) Award (with Raphael Cendrillon), and was a 1997 Laureate of the Belgium Royal Academy of Science. He received journal best paper awards from the IEEE Transactions on Signal Processing (with Geert Leus and with Daniele Giacobello) and from Elsevier Signal Processing (with Simon Doclo).
He was chairman of the IEEE Benelux Signal Processing Chapter (1998-2002), a member of the IEEE Signal Processing Society Technical Committee on Signal Processing for Communications, and President of EURASIP (European Association for Signal Processing, 2007-2008 and 2011-2012).
He has served as Editor-in-Chief for the EURASIP Journal on Applied Signal Processing (2003-2005), Area Editor for Feature Articles in IEEE Signal Processing Magazine (2012-2014), and has been a member of the editorial board of IEEE Transactions on Circuits and Systems II, IEEE Signal Processing Magazine, Integration-the VLSI Journal, EURASIP Journal on Wireless Communications and Networking, and Signal Processing. He is currently a member of the editorial board of EURASIP Journal on Advances in Signal Processing.