KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Randall Ali, Toon van Waterschoot, Marc Moonen, (2017), **A Noise Reduction Strategy for Hearing Devices Using an External Microphone.** **Technical Report 17-37**
Archived version	ftp://ftp.esat.kuleuven.be/stadius/rali/Reports/17-37.pdf
Author contact	randall.ali@esat.kuleuven.be *+ 32 (0)16 37 25 49*
IR	

(article begins on next page)

A Noise Reduction Strategy for Hearing Devices Using an External Microphone

Randall Ali, Toon van Waterschoot and Marc Moonen
KU Leuven, Dept. of Electrical Engineering (ESAT)
STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics
Kasteelpark Arenberg 10, 3001 Leuven, Belgium
Email: {randall.ali, toon.vanwaterschoot, marc.moonen}@esat.kuleuven.be

***Abstract*—In hearing devices such as hearing aids and cochlear implants, the use of an external microphone for audio processing in addition to the existing local microphone array, has proven to enhance the listening experience for a hearing-impaired user. However, as the processing has traditionally been accomplished by disabling the local microphones and amplifying the external microphone signal, this benefit comes at a cost of the loss of binaural cues and the undesired amplification of noise depending on the location of the external microphone. In this paper, we develop an extension to the ever popular Minimum Variance Distortionless Response (MVDR) beamformer that integrates the use of both the local microphone array and the external microphone, such that the external microphone can be gradually enabled or disabled as deemed relevant by the acoustic environment. We will demonstrate that this extension, referred to as the MVDR-XM, can provide a substantial noise reduction along with the local microphone array by defining an appropriate Relative Transfer Function (RTF) as well as converge to an MVDR with a predefined constraint using only the local microphone array. In cases where the external microphone signal is undesirable, reverting to the use of the local microphone array serves as a convenient feature of the MVDR-XM.**

I. Introduction

The presence of noise inevitably degrades speech intelligibility for individuals that suffer with a hearing impairment. Hearing devices such as Hearing Aids (HAs) and Cochlear Implants (CIs) are therefore commonly equipped with multiple microphones that need to perform a noise reduction task. As opposed to a single microphone, multiple microphones are able to exploit spatial diversity on top of spectral variations. This has generated an extensive development in the field of multi-microphone noise reduction and speech enhancement [1]. One popular algorithm in the area is the Minimum Variance Distortionless Response (MVDR) beamformer [2] [3]. In this algorithm, the received signal is preserved in a constraint direction to maintain a distortionless response and the noise variance is minimised in all other directions.

While the MVDR has proven to be an effective strategy, its design is usually limited in HAs and CIs due to the lack of physical space for the inclusion of additional components such as extra microphones. Consequently, there is an ongoing interest in the use of an external microphone for further improvement in speech enhancement. Existing systems that incorporate an external microphone with a communication link have already proven to provide benefits to HA and CI users [4]–[7]. Boothroyd [8], however, has mentioned that some subjects expressed concerns of persistent noise in very noisy environments as well as the problem of localisation. As some of these systems operate by disabling the local microphones on the hearing device and amplifying the external microphone signal, it is of no surprise that there will be issues with setting an appropriate gain, coupled with the loss of binaural cues.

State-of-the-art strategies for efficiently incorporating external microphone arrays within the context of wireless acoustic sensor networks have been described by Bertrand and Markovich-Golan [9] [10]. Szurley [11] has also pursued variants of the Multi-Channel Wiener Filter (MWF) [12] [13] so as to preserve the binaural cues when using an external microphone. An MWF is an appropriate strategy as no a-priori information is required and estimation is done from the second-order statistics of speech and noise signals. It can however result in unpredictable performance if these statistics are not accurately estimated.

In this paper, we focus on the use of one external microphone within the HA and CI context, and present an extension to a fixed constraint MVDR (MVDR-c) that integrates the use of both the local microphone array and the external microphone. Our formulation introduces a penalty term with a tuning parameter, β to the cost function of the MVDR-c. We also provide a method for redefining the fixed constraint that includes the external microphone along with the local microphone array. We subsequently refer to this MVDR strategy with an external microphone as the MVDR-XM. As will be proven, small values of β enable the information from the external microphone to be incorporated and significantly improve performance. Large values of β on the other hand, exclude the information from the external microphone and will converge to the performance of the MVDR-c that uses only the local microphone array. This serves as a useful feature of the MVDR-XM in cases where the external microphone may be in a very noisy environment or if there is a deficient communication link from the external microphone.

This paper is organised as follows. In section II, the data model and a review of the MVDR-c is provided. In section III, the proposed MVDR-XM strategy is presented. In section IV, simulations are done for comparing MVDR-c and MVDR-XM. In section V conclusions and future work will be summarised.

II. Data model and MVDR review

A. Microphone signals and general notation

We consider a noise reduction system that consists of a single microphone array of M microphones plus one additional external microphone, so that the total number of microphones is $M+1$. We also consider a scenario where there is only one desired speech signal in a noisy environment. Proceeding to formulate the problem in the short-time Fourier transform (STFT) domain, we can represent the received signal at one particular frequency ω and at time, t (dropping these indices for brevity), as:

$$\mathbf{y} = \mathbf{d}S + \mathbf{n} \tag{1}$$

where $\mathbf{y} = [\mathbf{y_a}\ \mathrm{y_e}]^T$, $\mathbf{y_a} = [\mathrm{y}_1\ \mathrm{y}_2\ \dots \mathrm{y_M}]^T$ are the local microphone signals in the array, $\mathrm{y_e}$ is the external microphone signal, $\mathbf{d} = [\mathbf{d_a}\ \mathrm{d_e}]^T$, is the Acoustic Transfer Function (ATF) from the speech source to all $M+1$ microphones and S, the speech signal. $\mathbf{n} = [\mathbf{n_a}\ \mathrm{n_e}]^T$ represents the noise contribution, which consists of a combination of correlated and uncorrelated noises. Variables with the subscript "$\mathbf{a}$" refer to the local microphone array and variables with the subscript "e" refer to the external microphone.

The spatial correlation matrices for all of the received signals, consisting of speech and noise as well as that of the noise only are given respectively as:

$$\mathbf{R_{yy}} = \mathbb{E}\{\mathbf{y}\mathbf{y}^H\} \tag{2}$$

$$\mathbf{R_{nn}} = \mathbb{E}\{\mathbf{n}\mathbf{n}^H\} \tag{3}$$

where $\mathbb{E}\{.\}$ is the expectation operator and ^{H} is the hermitian transpose. The spatial correlation matrices for the speech and noise and the noise only can also be calculated solely for the local microphone array signals respectively as $\mathbf{R_{y_a y_a}} = \mathbb{E}\{\mathbf{y_a}\mathbf{y_a}^H\}$ and $\mathbf{R_{n_a n_a}} = \mathbb{E}\{\mathbf{n_a}\mathbf{n_a}^H\}$. Given the non-stationary nature of speech and noise, exponential forgetting factors [14] can be used to calculate these correlation matrices.

The output signal, z, is then obtained through the linear filtering of the microphone signals, such that:

$$\mathrm{z} = \mathbf{w}^H\mathbf{y} \tag{4}$$

where $\mathbf{w} = [\mathbf{w_a}\ \mathrm{w_e}]^T$ is the complex-valued filter to be designed.

B. MVDR with a fixed constraint (MVDR-c)

The MVDR as proposed in [2] [3] minimises the total noise power (minimum variance), while preserving the received signal in a particular direction (distortionless response). Considering only the local microphone array, the problem can be formulated as follows:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \mathbf{w_a}^H\mathbf{R_{n_a n_a}}\mathbf{w_a} \\ \text{s.t.} \quad & \mathbf{w_a}^H\hat{\mathbf{d}}_\mathbf{a} = 1 \end{aligned} \tag{5}$$

where $\hat{\mathbf{d}}_\mathbf{a} = [\hat{\mathrm{d}}_{\mathrm{a},1}\ \hat{\mathrm{d}}_{\mathrm{a},2}\ \dots \hat{\mathrm{d}}_{\mathrm{a,M}}]^T$ is the assumed ATF from the speech source to the local microphone array that defines the constraint direction for which the speech is to be preserved. $\hat{\mathbf{d}}_\mathbf{a}$ may be based on a-priori assumptions regarding microphone characteristics, position, speaker location and room acoustics (e.g. no reverberation). For instance, it is not uncommon in hearing devices to assume that the speaker is directly in front of the user.

A common modification to (5) is to use a Relative Transfer Function (RTF), which is a normalisation of the constraint typically with the microphone that has the highest SNR. Choosing the first microphone as the reference in our case, we define the RTF, $\widetilde{\mathbf{d}}_\mathbf{a} = [1\ \frac{\hat{\mathrm{d}}_{\mathrm{a},2}}{\hat{\mathrm{d}}_{\mathrm{a},1}}\ \dots \frac{\hat{\mathrm{d}}_{\mathrm{a,M}}}{\hat{\mathrm{d}}_{\mathrm{a},1}}]^T$. In using the RTF, $\widetilde{\mathbf{d}}_\mathbf{a}$, instead of the ATF, $\hat{\mathbf{d}}_\mathbf{a}$ in (5), we effectively exclude the problem of de-reverberation and focus solely on noise reduction, as well as improve efficiency and robustness in the algorithm [15]. The optimal noise reduction filter for (5) that uses an RTF is then given by:

$$\mathbf{w_a} = \frac{\mathbf{R}^{-1}_{\mathbf{n_a n_a}}\widetilde{\mathbf{d}}_\mathbf{a}}{\widetilde{\mathbf{d}}^H_\mathbf{a}\mathbf{R}^{-1}_{\mathbf{n_a n_a}}\widetilde{\mathbf{d}}_\mathbf{a}} \tag{6}$$

which we will refer to as the MVDR-c.

The output is subsequently calculated as in (4), but only using the filters designed for the array, $\mathbf{w_a}$, from (6) and the microphone signals in the array, $\mathbf{y_a}$ (i.e. $\mathrm{w_e}$= 0). While this MVDR-c can provide a distortionless response if the speech truly lies in the direction defined by $\widetilde{\mathbf{d}}_\mathbf{a}$, in very noisy environments, speech intelligibility may still be affected if enough noise is not reduced. The incorporation of an external microphone provides an additional degree of freedom and if it is indeed close to the desired speaker, can improve the performance of the MVDR-c in such adverse acoustic environments. The following section outlines the strategy for using external microphone within the MVDR framework.

III. MVDR-XM

A. MVDR-XM Cost Function

While the external microphone can prove to be beneficial for noise reduction when it is close to the speaker, we cannot overlook the fact that the external microphone may not always be in such an ideal location. For instance, it is possible that the external microphone is close to a noise source or that there may be a deficient communication link from the external microphone, which in both cases can result in an undesirable performance. This motivates the need for a mechanism that can be used to softly include or exclude the contribution of the external microphone. We therefore proceed to modify the MVDR-c cost function of (5) with a tuning parameter, β (note that all $M+1$ microphone signals are now being used):

$$\begin{aligned} \min_{\mathbf{w}} \quad & \mathbf{w}^H\mathbf{R_{nn}}\mathbf{w} + \beta\left(\mathbf{w}^H\mathbf{e_e}\mathbf{e_e}^H\mathbf{w}\right) \\ \text{s.t.} \quad & \mathbf{w}^H\widetilde{\mathbf{d}} = 1 \end{aligned} \tag{7}$$

where $\mathbf{e_e} = [0\ 0\ \dots 0\ 1]^T$ is an $M+1$ dimensioned vector that is used to select the component corresponding to the external microphone signal, $\widetilde{\mathbf{d}} = [\widetilde{\mathbf{d}}_\mathbf{a}\ \widetilde{\mathrm{d}}_\mathrm{e}]^T$ is the RTF from the speech source to all $M+1$ microphones with the reference maintained as the first microphone of the array and $\widetilde{\mathrm{d}}_\mathrm{e} = \frac{\hat{\mathrm{d}}_\mathrm{e}}{\hat{\mathrm{d}}_{\mathrm{a},1}}$. In general,

$\widetilde{\mathrm{d}}_{\mathrm{e}}$ is unknown as $\hat{\mathrm{d}}_{\mathrm{e}}$, the part of the assumed ATF specifically from the speech source to external microphone, is subject to change. In the following section, a procedure is therefore given to obtain an estimate for the RTF $\widetilde{\mathrm{d}}_{\mathrm{e}}$.

The extra term in (7) is essentially a penalty term, which adds some value of $\beta \in [0\,\infty]$ to the bottom right entry of $\mathbf{R_{nn}}$. This entry of $\mathbf{R_{nn}}$ is the autocorrelation of the external microphone, hence β can be interpreted as an (inverse) weighting that controls how much we would like to incorporate the information from the external microphone. For instance, larger values of β will place a greater penalty on the external microphone, ultimately resulting in $\mathrm{w_e} \to 0$. As will be proven in section III-C, as $\beta \to \infty$, the optimal filter reduces to the MVDR-c in (6) that uses only the local microphone array.

B. RTF with the external microphone

In (7), $\widetilde{\mathbf{d}}$ can be fully estimated using the popular subtraction or decomposition methods for RTF estimation [16]. However, as we would like for the MVDR-XM to have a built in strategy of reverting to the MVDR-c when the external microphone is in an undesirable location, we proceed to maintain that $\widetilde{\mathbf{d}}_{\mathbf{a}}$ will be based on a-priori assumptions. Consequently, only an estimate for $\widetilde{\mathrm{d}}_{\mathrm{e}}$ is required as the position of the external microphone is subject to unpredicted changes.

One straightforward method to obtain an estimate to $\widetilde{\mathrm{d}}_{\mathrm{e}}$, which we denote as $\bar{\mathrm{d}}_{\mathrm{e}}$, is to perform a least squares estimation. Firstly, using only the local microphone array signals, the MVDR filter of (6) is applied to attain an estimate of the speech signal in the first microphone, $\tilde{S}_1 = \mathbf{w_a}^H \mathbf{y_a}$. The estimated RTF, $\bar{\mathrm{d}}_{\mathrm{e}}$, can then be found in a mean square error sense with the external microphone signal, $\mathrm{y_e}$, as follows:

$$\min_{\bar{\mathrm{d}}_{\mathrm{e}}} \quad \mathbb{E}\{|\bar{\mathrm{d}}_{\mathrm{e}}\,\tilde{S}_1 - \mathrm{y_e}|^2\} \tag{8}$$

with the solution:

$$\bar{\mathrm{d}}_{\mathrm{e}} = \mathbb{E}\left\{\frac{\mathrm{y_e}\,\tilde{S}_1^*}{\tilde{S}_1\,\tilde{S}_1^*}\right\} \tag{9}$$

As with the correlation matrices, $\bar{\mathrm{d}}_{\mathrm{e}}$ can also be calculated with exponential forgetting factors. $\bar{\mathrm{d}}_{\mathrm{e}}$ is subsequently substituted for $\widetilde{\mathrm{d}}_{\mathrm{e}}$ in (7).

C. MVDR-XM Solution and limiting cases

The cost function of (7) can be solved by defining a complex Lagrangian:

$$\begin{aligned}\mathcal{L} = \mathbf{w}^H\mathbf{R_{nn}}\mathbf{w} + \beta\,(\mathbf{w}^H\mathbf{e_e}\mathbf{e_e}^H\mathbf{w}) + \lambda(\mathbf{w}^H\widetilde{\mathbf{d}} - 1) \\ + \lambda^*(\widetilde{\mathbf{d}}^H\mathbf{w} - 1)\end{aligned} \tag{10}$$

where λ is a Lagrange multiplier. Setting the derivative of the Lagrangian with respect to $\mathbf{w}^H$ to 0 and imposing the constraint of $\mathbf{w}^H\widetilde{\mathbf{d}} = 1$ subsequently results in the optimal filter of a similar form to (6):

$$\mathbf{w}_{\text{mvdr-xm}} = \frac{\mathbf{R}_{\mathbf{nn}\beta}^{-1}\,\widetilde{\mathbf{d}}}{\widetilde{\mathbf{d}}^H\,\mathbf{R}_{\mathbf{nn}\beta}^{-1}\,\widetilde{\mathbf{d}}} \tag{11}$$

where $\mathbf{R_{nn}}_\beta = (\mathbf{R_{nn}} + \beta\mathbf{e_e}\mathbf{e_e}^H)$. The term $\beta\mathbf{e_e}\mathbf{e_e}^H$ is an $(M+1)$ x $(M+1)$ matrix, with β in the bottom right entry of the matrix and zeros elsewhere.

When $\beta = 0$, no restrictions are placed on the minimisation of the external microphone component and its information is incorporated into the MVDR solution. This suggests that smaller values of β should be selected if the estimated RTF, $\bar{\mathrm{d}}_{\mathrm{e}}$, is acceptable from (9) as well as if it is known that the external microphone is close to the speaker, i.e. being subject to good SNR conditions.

On the other hand, if the external microphone is not in a favourable environment or there is a deficient communication link from the external microphone, larger values of β will place a greater penalty on the external microphone so as to minimise its influence. In the limiting case of $\beta \to \infty$, it ultimately results in the exclusion of the external microphone from the optimal MVDR-XM filter in (11). To prove this, we can re-write the inverse in (11) by applying the matrix inversion lemma as $\beta \to \infty$:

$$\underset{\beta\to\infty}{\mathbf{R}_{\mathbf{nn}\beta}^{-1}} = \mathbf{R}_{\mathbf{nn}}^{-1} - \frac{\mathbf{R}_{\mathbf{nn}}^{-1}\mathbf{e_e}\mathbf{e_e}^H\mathbf{R}_{\mathbf{nn}}^{-1}}{\mathbf{e_e}^H\mathbf{R}_{\mathbf{nn}}^{-1}\mathbf{e_e}} \tag{12}$$

The noise correlation matrix, $\mathbf{R_{nn}}$ can also be re-written in block form:

$$\mathbf{R_{nn}} = \left[\begin{array}{c|c} \underset{(\mathrm{MxM})}{\mathbf{R}_{\mathbf{n_a n_a}}} & \underset{(\mathrm{Mx1})}{\mathbf{r_e}} \\ \hline \underset{(\mathrm{1xM})}{\mathbf{r_e}^H} & \underset{(\mathrm{1x1})}{\mathrm{r_{ee}}} \end{array}\right] \tag{13}$$

where the upper left block is the noise correlation matrix from the local microphone array, $\mathbf{r_e}$ is the noise cross correlation between the microphone array and the external microphone and $\mathrm{r_{ee}}$ is the noise autocorrelation of the external microphone. Applying a block inversion yields:

$$\mathbf{R}_{\mathbf{nn}}^{-1} = \left[\begin{array}{c|c} \mathbf{R}_{\mathbf{n_a n_a}}^{-1} + \frac{\mathbf{u_e}\mathbf{u_e}^H}{\mathrm{u_{ee}}} & \mathbf{u_e} \\ \hline \mathbf{u_e}^H & \mathrm{u_{ee}} \end{array}\right] \tag{14}$$

where $\mathbf{u_e} = -\mathbf{R}_{\mathbf{n_a n_a}}^{-1}\mathbf{r_e}(\mathrm{r_{ee}} - \mathbf{r_e}^H\mathbf{R}_{\mathbf{n_a n_a}}^{-1}\mathbf{r_e})$ and $\mathrm{u_{ee}} = (\mathrm{r_{ee}} - \mathbf{r_e}^H\mathbf{R}_{\mathbf{n_a n_a}}^{-1}\mathbf{r_e})^{-1}$. Substituting (14) into (12) subsequently results in:

$$\begin{aligned}\mathbf{R}_{\mathbf{nn}\beta}^{-1} &= \left[\begin{array}{c|c} \mathbf{R}_{\mathbf{n_a n_a}}^{-1} + \frac{\mathbf{u_e}\mathbf{u_e}^H}{\mathrm{u_{ee}}} & \mathbf{u_e} \\ \hline \mathbf{u_e}^H & \mathrm{u_{ee}} \end{array}\right] - \left[\begin{array}{c|c} \frac{\mathbf{u_e}\mathbf{u_e}^H}{\mathrm{y_{ee}}} & \mathbf{u_e} \\ \hline \mathbf{u_e}^H & \mathrm{u_{ee}} \end{array}\right] \\ &= \left[\begin{array}{c|c} \mathbf{R}_{\mathbf{n_a n_a}}^{-1} & \mathbf{0} \\ \hline \mathbf{0} & 0 \end{array}\right]\end{aligned} \tag{15}$$

Finally, it can be seen that the substitution of (15) into (11) results in (6), i.e. the MVDR-c. Hence, with the tuning parameter, β, the MVDR-XM optimal filter defined in (11) can provide a full exploitation of the external microphone to improve noise reduction performance, as well as ensure that there is a strategy of reverting to the MVDR-c defined only by the local microphone array as in (6).

IV. Simulations

The simulation environment consisted of a 0.5 second reverberant room with dimensions 7.1 m x 6.3 m x 5.2 m, a microphone array, one omnidirectional external microphone, a single speech source and a single localised noise source. The array consisted of 2 omnidirectional microphones with an inter-element spacing of 1 cm. For the speech source signal, six sentences separated by silence from the English Hearing-In-Noise Test (HINT) database [17] were used. This speech source was placed at the end-fire location of the array. $\widetilde{\mathbf{d}}_\mathbf{a}$ was defined as a steering vector based on this end-fire direction of arrival. The localised noise source signal was an excerpt of multitalker babble noise from Audiotec [18] and placed at the broadside direction of the array. Uncorrelated white noise was also added to each of the microphone signals such that the ratio of the speech signal power in the first microphone of the array to the uncorrelated white noise power was 13 dB. All simulations were performed using the Weighted Overlap and Add (WOLA) method [19], with an FFT size of 256, 50% overlap, and sampling frequency of 16 kHz. The room impulse responses were obtained using the randomised image method [20] and implemented from [21]. A perfect voice activity detector (VAD) was also used to obtain the relevant correlation matrices, with forgetting factors set to 0.995. The performance of the noise reduction strategies was evaluated using Speech Intelligibility-weighted SNR improvement (Δ SI-SNR) and Speech Intelligibility-weighted Spectral Distortion (SI-SD) measures defined in [13] (equations (45) - (47)), using the first microphone of the array (closer to the speech source) as the reference microphone.

In the simulations of the first column in figure 1, the external microphone was placed in the constraint (end-fire direction) at a distance of 20 cm away from the speech source in order to emulate a scenario where the speaker is wearing the external microphone. The Δ SI-SNR and SI-SD with respect to the first microphone of the array was then calculated as a function of the input SNR of this same reference microphone. The gain of the localised noise source was varied in order to achieve different input SNRs. As noted in figure 1 (a), although the input SNR at the first microphone degrades with increasing gain of the localised noise, the input SNR of the external microphone remains high (above 5dB) as the external microphone is indeed close to the speech source. Figures 1 (b) and 1 (c) illustrate the Δ SI-SNR and SI-SD respectively for the MVDR-c of section II-B and the MVDR-XM of section III for different values of β. In the simulations, these values of β have been normalised by the power of the external microphone signal, so for instance, $\beta = 10$ actually represents $\beta = 10 * |\mathrm{y_e}|^2$. As expected, for lower values of β, there is a large improvement in the Δ SI-SNR for the MVDR-XM as compared to that of the MVDR-c. This however, comes at the cost of an increased SI-SD. On the other end, it can be seen that for large values of β such as when $\beta = 1000$, the MVDR-XM converges to the performance of the MVDR-c. For values of β between the extremes, such as $\beta = 10$, a substantial Δ SI-SNR can still be achieved over the MVDR-c without having to compromise the SI-SD.

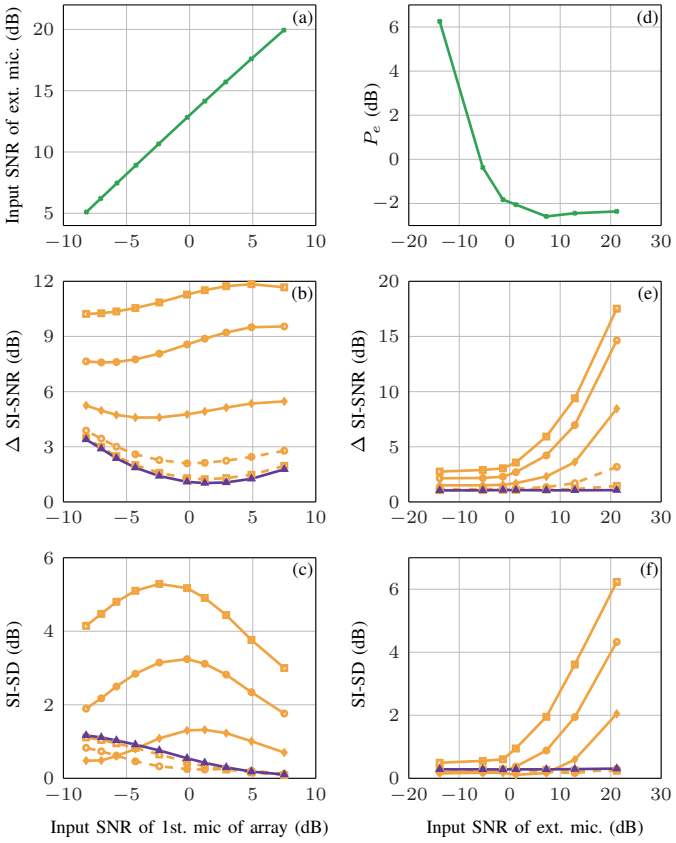


Fig. 1. The 1st column (subplots (a)-(c)) illustrates variations as a function of the input SNR of the 1st mic. in the array, with the external (ext.) mic. close to the speaker. (a) displays the corresponding changes in the input SNR of the ext. mic. and (b) and (c) compare the performance of the MVDR-c (—▲—) with the MVDR-XM for $\beta = 0$ (—■—), $\beta = 1$ (—●—), $\beta = 10$ (—◆—), $\beta = 100$ (- -●- -) and $\beta = 1000$ (- -■- -). In the 2nd column (subplots (d)-(f)), the ext. mic. was moved closer to the noise source while keeping the input SNR at the 1st mic. of the array at 0 dB. (d) demonstrates the degradation of the RTF estimate, $\bar{\mathrm{d}}_\mathrm{e}$ with decreasing input SNR of the ext. mic. (e) and (f) compare the performance of the MVDR-c with the MVDR-XM as a function of the input SNR of the ext. mic.

In the second column of figure 1, the input SNR at the first microphone of the array was fixed to 0 dB and the input SNR of the external microphone was varied by moving it closer to the localised noise source. In order to quantify the relative accuracy of the estimated RTF, $\bar{\mathrm{d}}_\mathrm{e}$, a ratio of the squared norm of the estimation error (with the assumed RTF, $\widetilde{\mathrm{d}}_\mathrm{e}$) to the squared norm of $\widetilde{\mathrm{d}}_\mathrm{e}$, was used:

$$P_e = 10\log_{10}\left\{\frac{||\widetilde{\mathrm{d}}_\mathrm{e} - \bar{\mathrm{d}}_\mathrm{e}||^2}{||\widetilde{\mathrm{d}}_\mathrm{e}||^2}\right\} \tag{16}$$

Figure 1 (d) displays this normalised estimation error, P_e, as a function of the input SNR of the external microphone. As the input SNR of the external microphone decreases, $\bar{\mathrm{d}}_\mathrm{e}$, clearly becomes more unreliable as indicated by the increase in P_e. Consequently, as seen in figures 1 (e) and (f), this results in a degradation of the performance of the MVDR-XM with

decreasing input SNR of the external microphone for lower values of β. Despite the poor RTF estimate, $\bar{\mathrm{d}}_{\mathrm{e}}$, at lower input SNR of the external microphone, there is still an improved performance over the MVDR-c case. This may be due simply to the fact that there is an extra degree of freedom with the external microphone. Nevertheless, it should not be expected that this will always be the case as performance is expected to degrade with even more unreliable $\bar{\mathrm{d}}_{\mathrm{e}}$. As in the previous simulations, it is also demonstrated that the MVDR-XM tends toward the the MVDR-c for larger values of β.

V. Conclusion

We have developed a strategy for the incorporation of an external microphone signal into a fixed-constraint MVDR beamformer (MVDR-c), which we refer to as the MVDR-XM. The MVDR-XM includes a tuning parameter, β, that facilitates the soft inclusion or exclusion of the external microphone signal. In scenarios where the external microphone is near to the desired speaker, lower values of β make this extra information available and the MVDR-XM is able to provide a significant benefit. In cases where the external microphone is in a very noisy environment or if there is a deficient communication link from the external microphone, larger values of β serve to inhibit the external microphone and the MVDR-XM converges to the MVDR-c. Although this paper does not discuss a binaural strategy, the fact that Relative Transfer Functions (RTFs) have been used in the MVDR-XM provides a framework for binaural cues to be preserved. In future work, we intend to explore this aspect further as well as variants to the MVDR-XM that can accommodate for RTF estimation as opposed to using fixed constraints. Furthermore, we also intend to implement this strategy within a Generalised Sidelobe Canceller (GSC) framework.

Acknowledgment

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 (OPTEC), IWT O&O Project nr. 150432 'Advances in Auditory Implants: Signal Processing and Clinical Aspects', KU Leuven Impulsfonds IMP/14/037, and KU Leuven Internal Funds VES/16/032. The scientific responsibility is assumed by its authors.

References

[1] S. Gannot, E. Vincent, S. Markovich-Golan, and A. Ozerov, "A consolidated perspective on multi-microphone speech enhancement and source separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, (to appear) 2017.

[2] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.

[3] E. Habets, J. Benesty, S. Gannot, and I. Cohen, *Speech Processing in Modern Communication: Challenges and Perspectives*. Berlin, Heidelberg: Springer, 2010, ch. 9: The MVDR Beamformer for Speech Enhancement, pp. 225–254.

[4] M. Ross, "FM Systems: A Little History and Some Personal Reflections," *Proceedings of the international conference, ACCESS: Achieving Clear Communication Employing Sound Solutions 2003 held November 2003 in Chicago, Illinois (USA), sponsored by Phonak*, pp. 17–27, 2003.

[5] K. L. Anderson, H. Goldstein, L. Colodzin, and F. Iglehart, "Benefit of S/N Enhancing Devices to Speech Perception of Children Listening in a Typical Classroom with Hearing Aids or a Cochlear Implant," *Journal of Educational Audiology*, vol. 12, pp. 16–30, 2005.

[6] E. M. Fitzpatrick, C. Séguin, D. R. Schramm, S. Armstrong, and J. Chénier, "The benefits of remote microphone technology for adults with cochlear implants." *Ear and hearing*, vol. 30, pp. 590–599, 2009.

[7] 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.

[8] A. Boothroyd, "Hearing aid accessories for adults: the remote FM microphone." *Ear and Hearing*, vol. 25, no. 1, pp. 22–33, 2004.

[9] A. Bertrand and M. Moonen, "Robust distributed noise reduction in hearing aids with external acoustic sensor nodes," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, 2009.

[10] S. Markovich-Golan, A. Bertrand, M. Moonen, and S. Gannot, "Optimal distributed minimum-variance beamforming approaches for speech enhancement in wireless acoustic sensor networks," *Signal Processing.*, vol. 107, no. C, pp. 4–20, Feb. 2015.

[11] J. Szurley, A. Bertrand, B. van Dijk, and M. Moonen, "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, 2016.

[12] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Transactions on Signal Processing*, vol. 50, no. 9, pp. 2230–2244, 2002.

[13] A. Spriet, M. Moonen, and J. Wouters, "Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction," *Signal Processing*, vol. 84, no. 12, pp. 2367–2387, 2004.

[14] S. Haykin, *Adaptive Filter Theory Fifth Edition*. Prentice Hall, 2013, ch. 7: Frequency-Domain and Subband Adaptive Filters.

[15] S. Gannot, D. Burshtein, and E. Weinstein, "Signal Enhancement Using Beamforming and Nonstationarity with Applications to Speech," *IEEE Transactions on Signal Processing*, vol. 49, no. 8, pp. 1614–1626, 2001.

[16] S. Markovich-Golan and S. Gannot, "Performance analysis of the covariance subtraction method for relative transfer function estimation and comparison to the covariance whitening method," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, South Brisbane, April 2015, pp. 544–548.

[17] M. Nilsson, S. D. Soli, and J. Sullivan, "Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise." *The Journal of the Acoustical Society of America*, vol. 95, no. 2, pp. 1085–1099, 1994.

[18] Auditec, "Auditory Tests (Revised), Compact Disc, Auditec, St. Louis," St. Louis, 1997.

[19] R. Crochiere, "A weighted overlap-add method of short-time Fourier analysis/Synthesis," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 1, pp. 99–102, 1980.

[20] E. De Sena, N. Antonello, M. Moonen, and T. van Waterschoot, "On the modeling of rectangular geometries in room acoustic simulations," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 4, pp. 774–786, April 2015.

[21] N. Antonello. (2016) Room impulse response generator with the randomized image method. [Online]. Available: https://github.com/nantonel/RIM.jl/tree/master/src/MATLAB