Citation/Reference	Randall Ali, Toon van Waterschoot, Marc Moonen, (2018), **Generalised Sidelobe Canceller for Noise Reduction in Hearing Devices using an external microphone**
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	
Journal homepage	https://2018.ieeeicassp.org/
Author contact	your email randall.ali@esat.kuleuven.be your phone number *+ 32 (0)16 37 25 49*
IR	

(article begins on next page)

GENERALISED SIDELOBE CANCELLER FOR NOISE REDUCTION IN HEARING DEVICES USING AN EXTERNAL MICROPHONE

Randall Ali[†], *Toon van Waterschoot*[†*] *and Marc Moonen*[†]

[†]KU Leuven, Dept. of Electrical Engineering (ESAT-STADIUS), Kasteelpark Arenberg 10, 3001 Leuven, Belgium

[*]KU Leuven, Dept. of Electrical Engineering (ESAT-ETC), e-Media Research Lab, Andreas Vesaliusstraat 13, 3000 Leuven, Belgium

ABSTRACT

The use of an external microphone in conjunction with an existing local microphone array can be particularly beneficial for noise reduction tasks that are critical for hearing devices, such as cochlear implants and hearing aids. Recent work has already demonstrated how an external microphone signal can be effectively incorporated into the common noise reduction technique of using a Minimum Variance Distortionless Response (MVDR) beamformer. In this paper, we provide a further extension, whereby an external microphone signal can be incorporated into an existing framework of a Generalised Sidelobe Canceller (GSC) that has been designed for a local microphone signal array. It will be shown that the resulting GSC with an external microphone results in an easily implementable addition to the existing GSC framework for a local microphone array, and can exhibit an improved noise reduction performance.

Index Terms— Noise Reduction, Speech Enhancement, Minimum Variance Distortionless Response, beamforming, Generalised Sidelobe Canceller, External Microphone

1. INTRODUCTION

As the presence of noise inevitably degrades speech intelligibility for individuals that suffer with a hearing impairment, a major design challenge for hearing devices such as hearing aids (HAs) and cochlear implants (CIs) is that of an effective noise reduction (NR) strategy. One popular NR strategy, known as the Minimum Variance Distortionless Response (MVDR) beamformer [1] [2], uses an array of microphones to focus on a desired signal from a particular direction, while attenuating signals from all other directions. A practical implementation of the MVDR beamformer is the Generalised Sidelobe Canceller (GSC) [3], which is an attractive solution for hearing devices due to its simplicity and lower complexity.

While the MVDR beamformer and GSC have proven to be effective NR strategies, their design is usually limited in hearing devices 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. In fact, existing systems that incorporate an external microphone with a communication link have already proven to provide benefits to HA and CI users [4–7].

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, KU Leuven C2-16-00449 'Distributed Digital Signal Processing for Ad-hoc Wireless Local Area Audio Networking', and KU Leuven Internal Funds VES/16/032. The scientific responsibility is assumed by its authors.

More recently, in the context of an MVDR beamformer, NR strategies for using the external microphone signal have been proposed in [8–10]. In this paper, we focus on the extension of an MVDR beamformer that uses an external microphone signal to its GSC counterpart. Specifically, we investigate how an external microphone signal can be incorporated into an existing GSC framework that has been designed for a local microphone array. In particular, we propose a strategy for incorporating the external microphone signal where the existing GSC framework is based on a priori assumptions.

This paper is organised as follows. The data model is provided in Section 2. A review of the MVDR and GSC for a local microphone array is given in Section 3. An MVDR that uses an external microphone signal and its GSC counterpart are presented in Section 4. Simulation results are discussed in Section 5 and conclusions are drawn in Section 6.

2. DATA MODEL

We consider a noise reduction system that consists of a single microphone array of M microphones plus one additional external microphone. 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, k, and one time frame, l, as:

$$\mathbf{y}(k,l) = \mathbf{h}(k,l)s_1(k,l) + \mathbf{n}(k,l) \tag{1}$$

where (dropping the dependency on k and l for brevity) $\mathbf{y} = [\mathbf{y_a}\ \mathrm{y_e}]^T$, $\mathbf{y_a} = [\mathrm{y_1}\ \mathrm{y_2}\ \ldots \mathrm{y_M}]^T$ are the signals from the local microphone array, $\mathrm{y_e}$ is the external microphone signal, $\mathbf{h} = [\mathbf{h_a}\ \mathrm{h_e}]^T$, is the Relative Transfer Function (RTF) among all $M+1$ microphones (i.e. a normalised Acoustic Transfer Function with respect to a reference (in this case, first) microphone in the local array), and s_1 is the speech component in the first microphone of the local array. $\mathbf{n} = [\mathbf{n_a}\ \mathrm{n_e}]^T$ represents the noise contribution, which consists of a combination of correlated and uncorrelated noise. Variables with the subscript "**a**" refer to the local microphone array and variables with the subscript "e" refer to the external microphone.

The $(M+1) \times (M+1)$ spatial correlation matrix for all of the received signals, consisting of speech and noise as well as that of the noise only is given respectively as:

$$\mathbf{R_{yy}} = \mathbb{E}\{\mathbf{y}\mathbf{y}^H\}; \quad \mathbf{R_{nn}} = \mathbb{E}\{\mathbf{n}\mathbf{n}^H\} \tag{2}$$

where $\mathbb{E}\{.\}$ is the expectation operator and ^{H} is the Hermitian transpose. The spatial correlation matrix 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 $\mathbf{R_{n_a n_a}} = \mathbb{E}\{\mathbf{n_a}\mathbf{n_a}^H\}$.

3. LOCAL MICROPHONE ARRAY PROCESSING

3.1. Minimum Variance Distortionless Response Beamformer

The MVDR as proposed in [1] [2] 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:

$$\begin{aligned} \min_{\mathbf{w_a}} \quad & \mathbf{w_a}^H \mathbf{R_{n_a n_a}} \mathbf{w_a} \\ \text{s.t.} \quad & \mathbf{w_a}^H \tilde{\mathbf{h}}_{\mathbf{a}} = 1 \end{aligned} \tag{3}$$

where $\tilde{\mathbf{h}}_{\mathbf{a}} = [\tilde{\mathrm{h}}_{\mathrm{a},1} \; \tilde{\mathrm{h}}_{\mathrm{a},2} \; \ldots \tilde{\mathrm{h}}_{\mathrm{a,M}}]^T$ is the a priori RTF for the local microphone array that defines the constraint direction for which the speech is to be preserved. $\tilde{\mathbf{h}}_{\mathbf{a}}$ can 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 the speaker location [11–13]. The optimal noise reduction filter for (3) is then given by:

$$\mathbf{w_a} = \frac{\mathbf{R}^{-1}_{\mathbf{n_a n_a}} \tilde{\mathbf{h}}_{\mathbf{a}}}{\tilde{\mathbf{h}}_{\mathbf{a}}^H \mathbf{R}^{-1}_{\mathbf{n_a n_a}} \tilde{\mathbf{h}}_{\mathbf{a}}} \tag{4}$$

which we will refer to as the MVDR-LM. The speech estimate, $\tilde{s}_a$, from this a priori MVDR is then obtained through the linear filtering of the microphone signals with the complex-valued filter $\mathbf{w_a}$:

$$\tilde{s}_a = \mathbf{w_a}^H \mathbf{y_a} \tag{5}$$

3.2. Generalised Sidelobe Canceller

In the practical implementation of the MVDR-LM as proposed by Griffiths and Jim [3], the constrained minimisation problem of (3) is converted into an unconstrained one. The resulting beamformer, known as the Generalised Sidelobe Canceller (GSC), is displayed in Figure 1, which consists of two branches. The top branch provides a speech reference by satisfying the constraint in (3) through the use of a fixed beamformer, $\mathbf{w_{fa}}$, i.e. $\mathbf{w_{fa}}^H \tilde{\mathbf{h}}_{\mathbf{a}} = 1$. The output of the top branch is then given by $\mathrm{y_f} = \mathbf{w_{fa}}^H \mathbf{y_a}$.

The bottom branch provides the noise reference signals $\mathbf{u_a} = [\mathrm{u}_1 \; \mathrm{u}_2 \ldots \mathrm{u}_{\mathrm{M}-1}]^T$ through the $M \times (M-1)$ blocking matrix, $\widetilde{\mathbf{C}}_{\mathbf{a}}$, which is defined as being orthogonal to the corresponding RTFs such that $\widetilde{\mathbf{C}}_{\mathbf{a}}^H \tilde{\mathbf{h}}_{\mathbf{a}} = \mathbf{0}$. Therefore, we can define $\widetilde{\mathbf{C}}_{\mathbf{a}}$ as follows:

$$\widetilde{\mathbf{C}}_{\mathbf{a}} = \begin{bmatrix} -\tilde{\mathrm{h}}^*_{\mathrm{a},2} & -\tilde{\mathrm{h}}^*_{\mathrm{a},3} & \ldots & -\tilde{\mathrm{h}}^*_{\mathrm{a,M}} \\ 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix} \tag{6}$$

where $\{.\}^*$ denotes the complex conjugate.

In solving the following unconstrained optimisation problem, the adaptive filter, $\mathbf{v_a}$, is adjusted such as to reduce the residual noise in the speech reference at each time frame, l[1]:

$$\min_{\mathbf{v_a}} \quad \mathbb{E}\{|\mathbf{w}^H_{\mathbf{fa}}(l)\mathbf{y_a}(l) - \mathbf{v}^H_{\mathbf{a}}(l)\widetilde{\mathbf{C}}^H_{\mathbf{a}}(l)\mathbf{y_a}(l)|^2\} \tag{7}$$

In order to avoid speech cancellation due to speech leakage into the noise reference, $\mathbf{v_a}$ is usually updated in frames where only noise

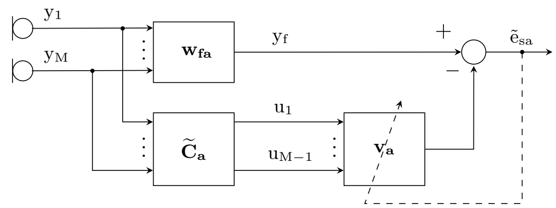

Fig. 1: Block scheme of the GSC-LM.

is present. In practice, the solution to (7) is often implemented with a Normalised Least Mean Squares (NLMS) approach [14] and the filter, $\mathbf{v_a}$, is consequently updated as:

$$\mathbf{v_a}(l+1) = \begin{cases} \mathbf{v_a}(l) + \frac{\mu}{||\mathbf{u_a}(l)||^2} \, \mathbf{u_a}(l) \, \tilde{\mathrm{e}}^*_{\mathrm{sa}}(l) & \text{noise frames} \\ \mathbf{v_a}(l) & \text{otherwise} \end{cases} \tag{8}$$

where $\mathbf{u_a}(l) = \widetilde{\mathbf{C}}^H_{\mathbf{a}}(l)\mathbf{y_a}(l)$, μ is the step size parameter for adaptation and

$$\tilde{\mathrm{e}}_{\mathrm{sa}}(l) = \mathrm{y_f}(l) - \mathbf{v}^H_{\mathbf{a}}(l)\widetilde{\mathbf{C}}^H_{\mathbf{a}}(l)\mathbf{y_a}(l) \tag{9}$$

is the output error corresponding to the speech estimate for the GSC-LM that converges to (5).

4. INCLUSION OF THE EXTERNAL MICROPHONE

4.1. Extension of the Minimum Variance Distortionless Response Beamformer

The MVDR-LM can be simply extended to incorporate the external microphone signal into what we refer to as the MVDR-XM:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \mathbf{w}^H \mathbf{R_{nn}} \mathbf{w} \\ \text{s.t.} \quad & \mathbf{w}^H \hat{\mathbf{h}} = 1 \end{aligned} \tag{10}$$

where $\hat{\mathbf{h}}$ is the RTF vector for the local microphones and external microphone to be estimated. Similarly to (4), the solution to (10) is given by:

$$\mathbf{w} = \frac{\mathbf{R}^{-1}_{\mathbf{nn}} \hat{\mathbf{h}}}{\hat{\mathbf{h}}^H \, \mathbf{R}^{-1}_{\mathbf{nn}} \, \hat{\mathbf{h}}} \tag{11}$$

The major concern with the MVDR-XM is with respect to the estimation of the RTF vector, $\hat{\mathbf{h}}$. Generally, the position of the external microphone is not known, and hence no a priori assumptions can be made about its location as could have been done for the MVDR-LM. Proceeding to represent $\hat{\mathbf{h}} = [\tilde{\mathbf{h}}_{\mathbf{a}} \mid \hat{\mathrm{h}}_{\mathrm{e}}]^T$, that partially consists of the a priori RTFs for the local microphone array, $\tilde{\mathbf{h}}_{\mathbf{a}}$, the problem is reduced to only requiring an estimate for the RTF from the external microphone to the first (reference) microphone in the array, $\hat{\mathrm{h}}_{\mathrm{e}}$, for the computation of (11).

One straightforward method to obtain $\hat{\mathrm{h}}_{\mathrm{e}}$ would be to perform a least squares estimation [9]. Using the estimate of the speech signal from the MVDR-LM in (5) in speech only frames, the estimated RTF, $\hat{\mathrm{h}}_{\mathrm{e}}$, can then be found in a mean square error sense with the external microphone signal, $\mathrm{y_e}$:

$$\min_{\hat{\mathrm{h}}_{\mathrm{e}}} \quad \mathbb{E}\{|\hat{\mathrm{h}}_{\mathrm{e}} \, \tilde{s}_a - \mathrm{y_e}|^2\} \tag{12}$$

[1]The dependency on l has been re-introduced in order to understand adaptation schemes. These quantities are still per frequency and the dependency on k will continue to be omitted for brevity.

with the solution:

$$\hat{\mathrm{h}}_{\mathrm{e}} = \frac{\mathbb{E}\{\mathrm{y_e}\,\tilde{s}_a^*\}}{\mathbb{E}\{\tilde{s}_a\,\tilde{s}_a^*\}} \tag{13}$$

Other methods for computing $\hat{\mathrm{h}}_{\mathrm{e}}$ are analysed in [10], however, we continue with the above procedure as it will be convenient for the associated GSC implementation.

4.2. Extension of the Generalised Sidelobe Canceller

The cost function for a GSC implementation of the MVDR-XM (GSC-XM) is a natural extension to (7) and is given by:

$$\min_{\mathbf{g}} \quad \mathbb{E}\{|\mathbf{w_f}^H(l)\mathbf{y}(l) - \mathbf{g}^H(l)\hat{\mathbf{C}}^H(l)\mathbf{y}(l)|^2\} \tag{14}$$

where $\mathbf{w_f} = [\mathbf{w_{fa}}\ \mathrm{w_{fe}}]^T$ with $\mathrm{w_{fe}}$ as the component of the fixed beamformer corresponding to the external microphone, $\mathbf{g} = [\mathbf{g_a}\ \mathrm{g_e}]^T$ is the adaptive filter to be designed, and the extended $(M+1) \times M$ blocking matrix is now given as:

$$\hat{\mathbf{C}} = \left[\begin{array}{cccc|c} -\tilde{\mathrm{h}}_{\mathrm{a},2}^* & -\tilde{\mathrm{h}}_{\mathrm{a},3}^* & \dots & -\tilde{\mathrm{h}}_{\mathrm{a},M}^* & -\hat{\mathrm{h}}_{\mathrm{e}}^* \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ \hline 0 & 0 & \dots & 0 & 1 \end{array}\right] \tag{15}$$

where the top left block of the matrix corresponds to $\widetilde{\mathbf{C}}_{\mathbf{a}}$ from (6).

The fixed beamformer, $\mathbf{w_f}$, can be readily simplified by setting $\mathrm{w_{fe}} = 0$. The role of the fixed beamformer within the context of a GSC is to satisfy the distortionless constraint, which can be accomplished regardless of the external microphone. As a result, $\hat{\mathrm{h}}_{\mathrm{e}}$ will only be required to define the blocking matrix, $\hat{\mathbf{C}}$, which requires an update for each time frame.

On substitution of (15) into (14), and with $\mathrm{w_{fe}} = 0$, the new speech estimate for the GSC-XM is given by:

$$\hat{\mathrm{e}}_{\mathrm{s}}(l) = \underbrace{\mathrm{y_f}(l) - \mathbf{g}_{\mathbf{a}}^H(l)\widetilde{\mathbf{C}}_{\mathbf{a}}^H(l)\mathbf{y_a}(l)}_{\text{speech reference, } \hat{\mathrm{e}}_{\mathrm{sa}}(l)} - \underbrace{\mathrm{g}_{\mathrm{e}}^*(l)(\mathrm{y_e}(l) - \hat{\mathrm{h}}_{\mathrm{e}}(l)\mathrm{y}_1(l))}_{\text{external mic. contribution, } \hat{\mathrm{e}}_{\mathrm{sx}}(l)} \tag{16}$$

which consists of two distinct components, $\hat{\mathrm{e}}_{\mathrm{sa}}$, a speech output that results from components related to those of the GSC-LM, and $\hat{\mathrm{e}}_{\mathrm{sx}}$, due to the incorporation of the external microphone signal. It is clear that when $\mathrm{g_e} = 0$, the contribution of the external microphone signal is disabled and the error or speech estimate will be identical to that of the GSC-LM in (9), i.e. $\mathbf{g_a} = \mathbf{v_a}$, and hence $\hat{\mathrm{e}}_{\mathrm{sa}} = \tilde{\mathrm{e}}_{\mathrm{sa}}$. However, in general, $\hat{\mathrm{e}}_{\mathrm{sa}} \neq \tilde{\mathrm{e}}_{\mathrm{sa}}$ as two different errors are minimised from the GSC-LM to the GSC-XM case. Nevertheless, $\hat{\mathrm{e}}_{\mathrm{sa}}$ still provides a speech reference as the local microphone signals pass through the fixed beamformer, $\mathbf{w_{fa}}$ and blocking matrix, $\widetilde{\mathbf{C}}_{\mathbf{a}}$. The difference between $\hat{\mathrm{e}}_{\mathrm{sa}}$ and $\tilde{\mathrm{e}}_{\mathrm{sa}}$ would then be due to a different filtering of the noise by $\mathbf{g_a}$ as opposed to $\mathbf{v_a}$.

With such a speech reference, $\hat{\mathrm{e}}_{\mathrm{sa}}$, $\hat{\mathrm{h}}_{\mathrm{e}}$ can now be conveniently calculated from (13), by replacing $\tilde{s}_a$ with $\hat{\mathrm{e}}_{\mathrm{sa}}$. This can be done in an adaptive manner as follows, which can then be substituted into

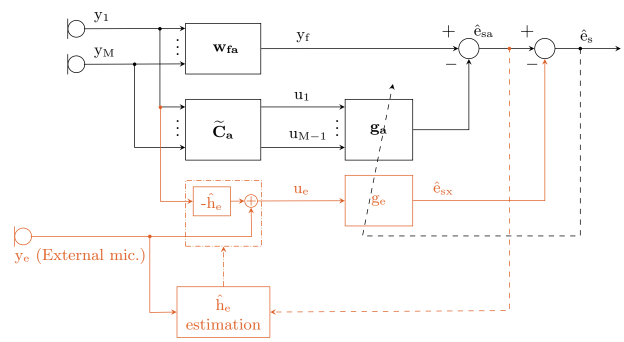

Fig. 2: (Colour online) Block scheme of the GSC-XM. The coloured sections represent the extensions to the GSC-LM for using the external microphone.

(15) to update the blocking matrix:

$$\hat{\mathrm{h}}_{\mathrm{e}}(l+1) = \begin{cases} \frac{r_{\mathrm{xa}}(l+1)}{r_{\mathrm{aa}}(l+1)} & \text{speech frames} \\ \hat{\mathrm{h}}_{\mathrm{e}}(l) & \text{otherwise} \end{cases} \tag{17}$$

where

$$r_{\mathrm{xa}}(l+1) = \alpha_e\, r_{\mathrm{xa}}(l) \;+\; (1-\alpha_e)\, \mathrm{y_e}(l)\, \hat{\mathrm{e}}_{\mathrm{sa}}^*(l) \tag{18}$$

$$r_{\mathrm{aa}}(l+1) = \alpha_e\, r_{\mathrm{aa}}(l) \;+\; (1-\alpha_e)\, \hat{\mathrm{e}}_{\mathrm{sa}}(l)\, \hat{\mathrm{e}}_{\mathrm{sa}}^*(l) \tag{19}$$

and $\alpha_e \in [0, 1]$ is a forgetting factor. With all the additional components for the GSC-XM defined, we can then proceed to perform an NLMS update of $\mathbf{g}$ in noise only frames such that:

$$\mathbf{g}(l+1) = \begin{cases} \mathbf{g}(l) + \frac{\mu}{||\mathbf{u}(l)||^2}\, \mathbf{u}(l)\ \hat{\mathrm{e}}_{\mathrm{s}}^*(l) & \text{noise frames} \\ \mathbf{g}(l) & \text{otherwise} \end{cases} \tag{20}$$

where $\mathbf{u}(l) = \hat{\mathbf{C}}^H(l)\mathbf{y}(l)$ and $\hat{\mathrm{e}}_{\mathrm{s}}(l)$ is computed from (16).

The GSC-XM procedure can be encapsulated by the block diagram as shown in Figure 2. The top branch remains unchanged from the GSC-LM because of the setting for $\mathrm{w_{fe}} = 0$, and hence only changes are made to the lower branch. $\widetilde{\mathbf{C}}_{\mathbf{a}}$ is now used along with $\hat{\mathrm{h}}_{\mathrm{e}}$ and the external microphone signal, $\mathrm{y_e}$, to create the new blocking matrix, $\hat{\mathbf{C}}$, and consequently an extended set of noise references, $\mathbf{u} = [\mathbf{u_a}\ \mathrm{u_e}]^T$, where $\mathrm{u_e}(l) = \mathrm{y_e}(l) - \hat{\mathrm{h}}_{\mathrm{e}}(l)\mathrm{y}_1(l)$. The speech reference, $\hat{\mathrm{e}}_{\mathrm{sa}}$, and $\mathrm{y_e}$ are used to calculate the estimate, $\hat{\mathrm{h}}_{\mathrm{e}}$, which updates the blocking matrix for the next time frame. $\mathbf{g_a}$ and $\mathrm{g_e}$ form the new adaptive filter, $\mathbf{g}$, which is updated by means of $\hat{\mathrm{e}}_{\mathrm{s}}$. The block diagram also intuitively depicts the two separate components of (16). A further advantage of such a block scheme is that it does not compromise the initial structure of the GSC-LM and can be interpreted as an "add-on" since it can easily be seen that if $\mathrm{g_e} = 0$, the GSC-XM is equivalent to the GSC-LM.

5. SIMULATIONS

In the simulations that follow, we considered a local microphone array with two omnidirectional elements separated by 1 cm, with an end-fire positioned speech source 1.3 m from the array, a broadside positioned localised noise source 1 m from the array, and a moving external microphone in a room of dimensions 6.9 m x 4.3 m x 2.6 m. For the speech source signal, seven sentences separated by silence from the English Hearing-In-Noise Test (HINT) database [15] were

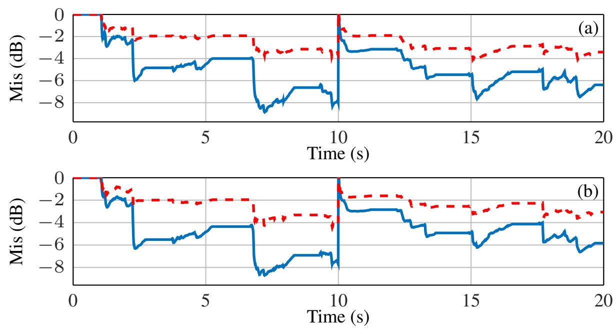

Fig. 3: (Colour online) Misalignment plots for (a) Real and (b) Imaginary parts for $\hat{\mathrm{h}}_{\mathrm{e}}$ in anechoic (——) and reverberant (- - -) conditions.

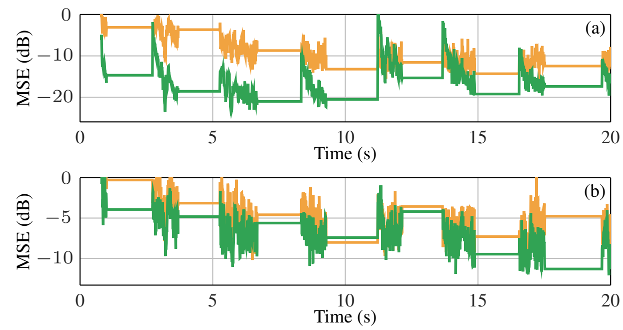

Fig. 4: (Colour online) MSE plots in (a) anechoic and (b) reverberant conditions for the GSC-LM (——) and GSC-XM (——).

used. The localised noise source signal was an excerpt of multitalker babble noise from Audiotec [16]. 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 30 dB. The input signal to noise ratio (SNR) at the first microphone of the array was approximately 4 dB.

All simulations were performed using the Weighted Overlap and Add (WOLA) method [17], with a Discrete Fourier Transform (DFT) size of 256, 50% overlap, and sampling frequency of 16 kHz. The room impulse responses were obtained using the randomised image method [18] and implemented from [19]. A perfect voice activity detector (VAD) was also used to access the signals in the relevant speech plus noise and noise only frames.

In the specific scenario that we analysed, the external microphone was initially placed 56 cm away from the speech source and after 10s it was moved closer to just 31 cm away from the speech source. The results were analysed in both anechoic and reverberant conditions (T60 = 300 ms). In both cases, the a priori RTF for the local microphone array, $\tilde{\mathbf{h}}_{\mathbf{a}}$, was defined using only the anechoic (direct) component of the room impulse response (RIR).

Figure 3 displays the misalignment for both real and imaginary parts between the estimated RTF for the external microphone, $\hat{\mathrm{h}}_{\mathrm{e}}$, and the anechoic RTF for the external microphone, which we denote as $\tilde{\mathrm{h}}_{\mathrm{e}}$. Since $\tilde{\mathbf{h}}_{\mathbf{a}}$ is defined as the anechoic part of the RIR, $\tilde{\mathrm{h}}_{\mathrm{e}}$ is the corresponding a priori RTF to which $\hat{\mathrm{h}}_{\mathrm{e}}$ will attempt to estimate. The misalignment was calculated in each time frame for all K frequency bins (for the real and imaginary parts accordingly) as:

$$\text{Mis (dB)} = 10\log_{10}\frac{\sum_{k=1}^{K}|\tilde{\mathrm{h}}_{\mathrm{e}}(k) - \hat{\mathrm{h}}_{\mathrm{e}}(k)|^2}{\sum_{k=1}^{K}|\tilde{\mathrm{h}}_{\mathrm{e}}(k)|^2} \tag{21}$$

It can be observed that in both anechoic and reverberant cases, $\hat{\mathrm{h}}_{\mathrm{e}}$ is able to adapt to changes in the position of the external microphone. As expected, the convergence for the anechoic case is better than that for the reverberant due to the definition of the a priori RTF. With reverberation, $\widetilde{\mathbf{C}}_{\mathbf{a}}$ is no longer able to fully block the speech component and there is inevitable speech leakage, which will consequently degrade the speech reference, $\hat{\mathrm{e}}_{\mathrm{sa}}$, for the updating of $\hat{\mathrm{h}}_{\mathrm{e}}$. Nevertheless, as there is still a degree of convergence in the reverberant case, it suggests that $\hat{\mathrm{e}}_{\mathrm{sa}}$ can still be a relevant choice for updating $\hat{\mathrm{h}}_{\mathrm{e}}$. In fact, this demonstrates that the performance is limited by the choice for the a priori models used for the local microphone array as opposed to the additional estimation procedures which have been proposed in the GSC-XM.

Figure 4 displays the resulting mean squared error (MSE) from (7) and (14), which has been calculated in noise only frames. The impact of the convergence of $\hat{\mathrm{h}}_{\mathrm{e}}$ is now evident. In the anechoic case (Figure 4 (a)), it can be observed that the GSC-XM is able to attain a lower MSE than that of the GSC-LM and adapts to changes in the external microphone position accordingly. In the reverberant case (Figure 4 (b)), the MSE for the GSC-XM is still lower than that of the GSC-LM, but not to the extent as that for the anechoic case, due to the convergence of $\hat{\mathrm{h}}_{\mathrm{e}}$. It should also be noted that both the GSC-XM and GSC-LM have a higher MSE as opposed to the anechoic conditions, indicative of the a priori assumptions not being satisfied. Regardless, these results suggest that a GSC-XM proposed in this paper is indeed an improvement upon a GSC-LM based on a priori assumptions. Furthermore, it is also shown that neither the performance or structure of the GSC-LM needs to be compromised in order to facilitate the external microphone addition. The corresponding audio files from these simulations can be heard at [20].

6. CONCLUSION

A strategy for incorporating the external microphone signal into an existing GSC framework based on a priori assumptions has been developed. This strategy, referred to as the GSC-XM, requires a minor extension to the blocking matrix and an RTF estimation procedure, neither of which compromise the existing GSC framework for the local microphone array. Simulation results have also demonstrated that the GSC-XM is able to adapt to changes in the position of the external microphone and can provide an improved performance over the GSC that uses only the local microphone array.

7. REFERENCES

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

[2] 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.

[3] L. Griffiths and C. Jim, "An alternative approach to linearly

constrained adaptive beamforming," *IEEE Transactions on Antennas and Propagation*, vol. 30, no. 1, pp. 27–34, 1982.

[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] 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.

[9] R. Ali, T. van Waterschoot, and M. Moonen, "A noise reduction strategy for hearing devices using an external microphone," 2017, ESAT-STADIUS Technical Report TR 17-37, KU Leuven, Belgium.

[10] N. Gößling, D. Marquardt, and S. Doclo, "Performance analysis of the extended binaural MVDR beamformer with partial noise estimation in a homogeneous noise field," *2017 Hands-free Speech Communications and Microphone Arrays (HSCMA)*, vol. San Francisco, pp. 1–5, 2017.

[11] J. E. Greenberg and P. M. Zurek, "Evaluation of an adaptive beamforming method for hearing aids," *Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1662–1676, 1992.

[12] J. M. Kates and M. R. Weiss, "A comparison of hearing-aid array-processing techniques," *Journal of the Acoustical Society of America*, vol. 99, no. 5, pp. 3138–3148, 1996.

[13] A. Spriet, L. Van Deun, K. Eftaxiadis, J. Laneau, M. Moonen, B. van Dijk, A. van Wieringen, and J. Wouters, "Speech understanding in background noise with the two-microphone adaptive beamformer BEAM in the Nucleus Freedom Cochlear Implant System." *Ear and hearing*, vol. 28, no. 1, pp. 62–72, 2007.

[14] S. Haykin, *Adaptive Filter Theory Fifth Edition*. Prentice Hall, 2013.

[15] 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.

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

[17] 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.

[18] 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.

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

[20] (2017). [Online]. Available: ftp://ftp.esat.kuleuven.be/stadius/rali/Reports/ICASSP%202018/Audio%20Files