RTF Based LCMV Beamformer with Multiple Reference Microphones

Amos Schreibman, Anna Barnov, Alex Gendelman, Eli Tzirkel

Speech and Acoustic Advanced Technologies, General Motors
Email: {amos.schreibman, anna.barnov, alex.gendelman, eli.tzirkel}@gm.com

***Abstract*—Microphone arrays are widely used for speech enhancement applications. We consider the enhancement of multiple desired speakers in a noisy environment while utilizing the Relative Transfer Function (RTF) approach for the beamformer realization. When beamforming operation relies on an estimated RTF, it produces a signal with reduced noise compared to the signal received by some reference microphone associated with the RTF, while maintaining its speech component undistorted. The reference microphone is usually chosen as the microphone with the highest Signal-to-Noise Ratio (SNR). For large arrays, in a multiple desired speaker environment, there is no single reference microphone which maximizes SNR for all desired speakers. The commonly used RTF technique may result in compromised performance for one or more desired speakers. Here, we propose an alternative scheme that considers multiple reference microphones. Each reference microphone is associated with a single desired speaker such that it maximizes its input SNR, in comparison to other array microphones. We show that by this technique, the beamformer maintains the desired noise reduction without compromising SNR for any of the desired speakers. We present analytical analysis for a 2 microphone array in free-field propagation for far-field sources. Based on this analysis some design criteria are derived and finally evaluated both in synthetic and recorded room environments.**

I. Introduction

New arising applications in the fields of speech recognition, VOIP and cellular communication present new challenges for speech enhancement algorithms. Targeting applications in the domains of smart home, fitness and medicine, we encounter severe noise conditions in highly dynamic environments. As technology advances and deployment of microphone arrays becomes common, multichannel processing, and beamforming techniques in particular, are utilized.

Commonly used beamforming designs optimize the Minimum Variance Distortionless Response (MVDR) criterion [1], which minimizes the noise power at the output, while maintaining the desired speech signal undistorted. Many audio capturing applications, such as teleconferencing, surveillance and broadcast, require transmission of multiple desired speech sources simultaneously. For these applications, the multiple constraints extension of the MVDR, known as the Linearly Constrained Minimum Variance (LCMV) beamformer, is used [2], [3].

For a beamformer design, the desired source Acoustic Transfer Function (ATF) is required. It reflects the source spatial properties, as its position, and characterizes the acoustic setup with no dependency on the emitted signal. Estimation of the ATF may be a complex task as it requires the emitted signal as reference, which in many real life scenarios may not be applicable. It has been proposed to estimate the Relative Transfer Function (RTF) instead [4]. The RTF technique relies only on the observed signals received by the microphones, thus applicable in practice. Different RTF estimation techniques are available and are shown to perform well in various noise environments and various array geometries [5]–[7]. However, all RTF estimation techniques require a choice of a reference microphone. The reference microphone is usually chosen as the microphone with the best input Signal-to-Noise Ratio (SNR) [8].

When beamforming operation relies on an estimated RTF, it produces a signal with reduced noise compared to the signal received by the reference microphone, while maintaining its speech component undistorted. In a multiple desired speakers scenario, for distributed microphone arrays systems or microphone arrays with large apertures, the traditional approach of using a single reference microphone for all desired RTFs may result in reduced output SNR for one or more desired speakers. This occurs when one or more desired speakers are distant from the chosen reference microphone. In the described scenario, the distant sources will maintain the power as received by the reference microphone, instead of producing a higher level signal by choosing the nearest microphone as reference.

In the following, we describe an alternative scheme that considers multiple different reference microphones for the estimated RTFs, instead of a single, shared microphone. We denote it by the Multi Reference LCMV (MR-LCMV). Each reference microphone is associated with a single desired speaker such that it maximizes its input SNR. We show that by this scheme the beamforming reduces noise without compromising the output speech levels for anyone of the desired speakers.

We present analytical analysis for a two microphone array in free-field for far-field sources, based on which, design guidelines for the MR-LCMV beamformer are derived. The proposed MR-LCMV is then analyzed in a simulated room environment as well as in office environment relying on real recorded data. It is evaluated in comparison to the traditional LCMV, considering SNR improvement, distortion and White Noise Gain (WNG) [9].

The paper is organized as follows. The multiple desired speakers problem is formulated in Section II and the traditional LCMV, which utilizes the RTF techniques, is described in Section III. The proposed MR-LCMV is presented and analyzed in Section IV. It is evaluated and compared to the traditional LCMV in Section V.

II. Problem Formulation and notation

Consider S desired speech signals propagating in a reverberant enclosure on an M element microphone array. The received signal is contaminated by additive noise comprising any combination of a point source, diffuse and spatially white noises.

Let

$$\mathbf{x}(\ell,k)=\left[\begin{array}{cccc} x_1(\ell,k) & x_2(\ell,k) & \cdots & x_M(\ell,k)\end{array}\right]^T \quad (1)$$

be the M dimensional Short Time Fourier Transform (STFT) domain representation of the stacked received microphones signals vector in time-frame ℓ and frequency bin k. The operator $(\cdot)^T$ denotes transpose.

The received signal $\mathbf{x}(\ell,k)$ can be decomposed into:

$$\mathbf{x}(\ell,k)=\sum_{i=1}^{S}\mathbf{h}_i(k)s_i(\ell,k)+\mathbf{v}(\ell,k), \quad (2)$$

where $s_i(\ell,k)$ is the speech signal of the ith speaker, $\mathbf{h}_i(k)$ is the corresponding ATF vector to each of the array microphones and $\mathbf{v}(\ell,k)$ is the noise vector, for $i=1,\ldots,S$.

The problem at hand is to design a beamformer $\mathbf{w}(\ell,k)$, such that the output signal

$$y(\ell,k)=\mathbf{w}^H(\ell,k)\mathbf{x}(\ell,k), \quad (3)$$

with operator $(\cdot)^H$ denoting conjugate-transpose, maximizes SNR for all S desired sources, while maintaining distortionless response for all. In the following, time-frame and frequency bin indices are omitted for brevity.

III. Traditional Solution and Problem Statement

A commonly used technique to address the problem of multiple desired speakers enhancement in noisy environment is the LCMV beamformer [1], defined as:

$$\mathbf{w}_{\text{LCMV}}=\underset{\mathbf{w}}{\operatorname{argmin}}\,\mathbf{w}^H\mathbf{\Phi}_{\mathbf{v}}\mathbf{w} \quad \text{s.t.} \quad \mathbf{C}^H\mathbf{w}=\mathbf{g} \quad (4)$$

where

$$\mathbf{C}=\left[\begin{array}{cccc}\tilde{\mathbf{h}}_1 & \tilde{\mathbf{h}}_2 & \cdots & \tilde{\mathbf{h}}_S\end{array}\right] \quad (5)$$

is the $M\times S$ constraints matrix consisting of the desired sources RTFs [4], such that $\tilde{\mathbf{h}}_i=\frac{\mathbf{h}_i}{h_r}$ for $i=1,\ldots,S$ and $r\in\{1,\ldots,M\}$ is the index of the reference microphone, traditionally chosen as 1 or as the microphone with the best input SNR. The vector $\mathbf{g}=\left[\begin{array}{cccc}1 & 1 & \cdots & 1\end{array}\right]^T$, of length S, is the corresponding constraints vector and $\mathbf{\Phi}_{\mathbf{v}}$ is the spatial noise coherence matrix.

The well known closed form solution to (4) is given by [1]:

$$\mathbf{w}_{\text{LCMV}}=\mathbf{\Phi}_{\mathbf{v}}^{-1}\mathbf{C}\left(\mathbf{C}^H\mathbf{\Phi}_{\mathbf{v}}^{-1}\mathbf{C}\right)^{-1}\mathbf{g}. \quad (6)$$

The RTF based LCMV beamformer maintains distortionless response towards the desired speech components as received by the reference microphone. When considering distributed microphone array systems or arrays with large apertures, the signal received by the reference microphone might not be optimal for all desired speakers in terms of SNR, that is, with reduced power of the speech component. This is expected, for example, in a diffuse noise scenario, for near field sources, when one of the desired speakers is positioned closest to some microphone $m\in\{1,\ldots,M\}\setminus\{r\}$. This results in reduced SNR for that speaker at the beamformer output.

IV. Proposed Solution

To maximize the output SNR for all desired speakers, while exploiting the important properties of RTFs and RTF based beamformers [7], we propose the use of multiple reference microphone, such that $\tilde{\mathbf{h}}_i$ in (5) becomes:

$$\tilde{\mathbf{h}}_i=\frac{\mathbf{h}_i}{h_{r_i}}, \quad (7)$$

where $r_i\in\{1,\ldots,M\}$ is the index of the microphone with the maximal input SNR for speaker i. The proposed MR-LCMV approach assures that the power of all desired speech components is maximized at the output of the beamformer.

In this section we provide analytical analysis of the MR-LCMV WNG in the simple case of $M=2$ microphones for $S=2$ far-field sources in free-field propagation and spatial white noise, that is $\mathbf{\Phi}_{\mathbf{v}}=\mathbf{I}$, where $\mathbf{I}$ is the identity matrix of dimension M.

While this scenario doesn't require the multi reference approach, as all desired speech signals are impinging with the same power on all microphones, as will be shown here, it will drive guidelines for the design criteria of the proposed MR-LCMV. Those will be verified in next sections in near field reverberant enclosure.

A. Problem Formulation for Analytical Analysis

Consider the 2 sources impinging on the 2 element microphone array from angles θ_1 and θ_2. Array interspacing is d meters, as depicted in Fig. 1.

The ATF for some frequency bin k, sampling frequency f_s and STFT resolution K, is given by:

$$\mathbf{h}_i=\left[\begin{array}{c}\exp(-j2\pi\frac{kf_s}{K}\tau_i)\\ \exp\left(-j2\pi\left(\frac{kf_s}{K}\tau_i+\frac{d}{\lambda_k}\cos\left(\theta_i\right)\right)\right)\end{array}\right], \quad (8)$$

where τ_i is the propagation time from source i to the first microphone, $\lambda_k=\frac{Kc}{kf_s}$ is wavelength corresponding to frequency bin k and c is the velocity of sound.

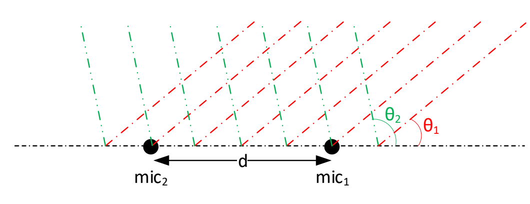


Fig. 1: 2 elements microphone array in free-field configuration with 2 far-field sources.

Without loss of generality, we choose $r=1$ for the traditional LCMV beamformer and for the MR-LCMV, $r_1=1$

and $r_2 = 2$. Considering (8), the LCMV and MR-LCMV constraints matrices are

$$\mathbf{C}_{\text{LCMV}} = \begin{bmatrix} 1 & 1 \\ \exp(-j\phi_1) & \exp(-j\phi_2) \end{bmatrix} \quad (9)$$

and

$$\mathbf{C}_{\text{MR-LCMV}} = \begin{bmatrix} 1 & \exp(+j\phi_2) \\ \exp(-j\phi_1) & 1 \end{bmatrix}, \quad (10)$$

respectively, with $\phi_i = \frac{2\pi d}{\lambda_k} \cos(\theta_i)$.

Substituting 9 into 6, with $\mathbf{g} = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$ and $\mathbf{\Phi_v} = \mathbf{I}$, yields:

$$\mathbf{w}_{\text{LCMV}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad (11)$$

with $\|\mathbf{w}_{\text{LCMV}}\|^2 = 1$.

This is an expected result suggesting that the beamformer output is in fact the reference signal x_1. This, since the problem imposes 2 constraints on a 2 microphone system with the reference signal meeting both constraints, leaving no degrees of freedom for noise minimization.

For MR-LCMV, the solution is not trivial. Neither of the 2 received signals, x_1 nor x_2, meets both distortionless response constraints. Substituting 10 into 6, with some calculations, yields:

$$\mathbf{w}_{\text{MR-LCMV}} = \frac{1}{\Delta} \begin{bmatrix} 1 - \exp(+j\phi_1) + \exp(+j\phi_2) - \exp(-j\psi) \\ 1 + \exp(-j\phi_1) - \exp(-j\phi_2) - \exp(-j\psi) \end{bmatrix} \quad (12)$$

where $\psi \triangleq \phi_1 - \phi_2$ and $\Delta = 4\sin^2(\psi/2)$.

Unlike the LCMV case, for the MR-LCMV, the impact on white noise is not trivial. The analysis of the MR-LCMV WNG is provided next in Sec. IV-B.

B. WNG Analysis

The metric of a beamformer WNG, or equivalently the beamformer norm, indicates its robustness and is many times considered as an additional design criterion [10]. In the scenario discussed here, that is for $\mathbf{\Phi_v} = \mathbf{I}$, the WNG analysis is three fold: 1) Robustness analysis of the MR-LCMV. 2) Analysis of the maximal gain that can be achieved for spatial white noise (such as sensors noise). 3) Analysis of the robustness upper bound of the MR-LCMV for any $\mathbf{\Phi_v}$.

By applying trigonometry identities, the norm of (12) can be shown to be:

$$\|\mathbf{w}_{\text{MR-LCMV}}\|^2 = \frac{1}{\sin^2(\frac{\psi}{2})} \left[1 - \cos(\frac{\psi}{2}) \cos(\phi_2 + \frac{\psi}{2}) \right] \quad (13)$$

and is depicted in Fig. 2 for $d/\lambda = 0.28$ for all θ_1 and θ_2 combinations.

From (13), interesting observations can be noted:

B.1 Close Direction Of Arrivals (DOAs) may cause an increased beamformer norm:

$$\lim_{\psi \to 0, \phi_2 \neq 0} \|\mathbf{w}_{\text{MR-LCMV}}\|^2 \to \infty. \quad (14)$$

B.2 Broadside position for one of the sources induces unity norm:

$$\lim_{\phi_2 \to 0, \psi \neq 0} \|\mathbf{w}_{\text{MR-LCMV}}\|^2 \to 1. \quad (15)$$

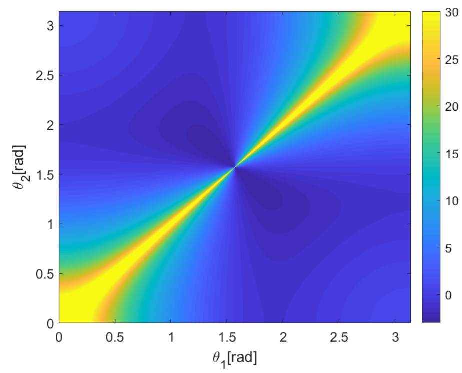


Fig. 2: $\|\mathbf{w}_{\text{MR-LCMV}}\|^2$ in dB, for $d/\lambda = 0.28$ and various θ_1, θ_2.

B.3 Observing Fig. 2 it is clear that $\|\mathbf{w}_{\text{MR-LCMV}}\|^2$ depends on θ_1 and θ_2, with its minimum obtained when $\theta_1 = \pi - \theta_2$, that is, when the two sources are positioned symmetrically with respect to the array center.

Considering B.3, by substituting $\theta_1 = \pi - \theta_2$, or alternatively $\psi = -2\phi_2$, into (13) it can be shown that in that special case $\|\mathbf{w}_{\text{MR-LCMV}}\|^2 \leq \|\mathbf{w}_{\text{LCMV}}\|^2 = 1$ for

$$|cos(\theta_1)| = |cos(\theta_2)| \leq \frac{1}{4}\frac{\lambda}{d}, \quad (16)$$

which implies that for $d/\lambda \leq 0.25$, $\|\mathbf{w}_{\text{MR-LCMV}}\|^2 \leq 1$ $\forall \{\theta_1, \theta_2 | \theta_1 = \pi - \theta_2\}$.

This is an interesting observation, as it suggests that in the discussed scenario, the MR-LCMV obtains some spatial white noise minimization even for the fully determined constraints linear system.

A more important observation is that, generally, the MR-LCMV norm is not bounded, unlike that of the traditional LCMV. This must be considered in the MR-LCMV design criterion, as discussed next.

C. MR-LCMV Design Criterion

The traditional LCMV beamformer is designed according to (4). In the scenario discussed in this sequel, it minimizes the noise power output, while maintaining distortionless response towards all considered sources. To consider numerical robustness, additional constraint on the beamformer norm is introduced, such that (4) becomes:

$$\mathbf{w}_{\text{LCMV}} = \underset{\mathbf{w}}{\operatorname{argmin}}\, \mathbf{w}^H \mathbf{\Phi_v} \mathbf{w} \quad \text{s.t.} \quad \mathbf{C}^H \mathbf{w} = \mathbf{g}, \quad \|\mathbf{w}\|^2 \leq \delta^2 \quad (17)$$

where feasible values for δ^2 must satisfy [10]:

$$\delta^2 \geq \|\mathbf{q}\|^2, \quad (18)$$

for

$$\mathbf{q} = \underset{\mathbf{w}}{\operatorname{argmin}}\, \mathbf{w}^H \mathbf{w} \quad \text{s.t.} \quad \mathbf{C}^H \mathbf{w} = \mathbf{g}. \quad (19)$$

Note that for the problem discussed in this section, that is $M = 2$, $S = 2$, $\mathbf{g} = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$ and $\mathbf{\Phi_v} = \mathbf{I}$, $\|\mathbf{q}\|^2$ coincides with (13). The fact that (13) is not bounded suggests that for the MR-LCMV beamformer it is possible that no feasible values for δ^2 exist and hence the norm cannot be controlled and very high norm values may be experienced.

To resolve this, we propose to adopt the soft constraints approach proposed in [11]. There, weighting between the minimization term and distortionless constraint is applied to allow better noise reduction at the expense of a perfect distortionless response. Considering that, we propose the design criterion of the MR-LCMV beamformer to be:

$$\mathbf{w}_{\text{MR-LCMV}} = \underset{\mathbf{w}}{\operatorname{argmin}}\, \mathbf{w}^H \mathbf{\Phi}_{\mathbf{v}} \mathbf{w} + \mu |\mathbf{C}^H_{\text{MR-LCMV}} \mathbf{w} - \mathbf{g}|^2 \quad \text{s.t.} \quad \|\mathbf{w}\|^2 \leq \delta^2, \tag{20}$$

where $\mu \geq 0$ is a weighting factor. Note that for $\mu = \infty$, (20) becomes (17).

To choose the value for μ, consider:

$$\mathbf{q}_{\text{MR-LCMV}}(\mu) = \underset{\mathbf{w}}{\operatorname{argmin}}\, \mathbf{w}^H \mathbf{w} + \mu |\mathbf{C}^H_{\text{MR-LCMV}} \mathbf{w} - \mathbf{g}|^2, \tag{21}$$

with the closed form solution:

$$\mathbf{q}_{\text{MR-LCMV}}(\mu) = \mu \left(\mathbf{I} + \mu \mathbf{C}_{\text{MR-LCMV}} \mathbf{C}^H_{\text{MR-LCMV}}\right)^{-1} \mathbf{C}_{\text{MR-LCMV}} \mathbf{g}. \tag{22}$$

Note that $\|\mathbf{q}_{\text{MR-LCMV}}(\mu)\|^2 \leq \|\mathbf{w}_{\text{MR-LCMV}}\|^2 \; \forall \delta^2$ and acts as the lower bound for δ^2.

Also, noting that $\|\mathbf{q}_{\text{MR-LCMV}}(\mu)\|^2$ monotonically increases with μ, μ can be obtained as $\mu = \mu_{\max}$ by solving

$$\|\mathbf{q}_{\text{MR-LCMV}}(\mu_{\max})\|^2 = \beta^2 \tag{23}$$

for $0 \leq \beta^2 \leq \delta^2$.

V. Experimental Results

In this section we evaluate the proposed MR-LCMV method and compare it with the traditional LCMV. The section is divided into two parts as follows: In section V-A we simulate an Audio Visual (AV) conference room, and compare between the traditional LCMV and the proposed MR-LCMV under spatially white and diffuse noise fields, for two desired speakers. In section V-B, real room recordings are taken for two real speakers and the two approaches are compared.

A. Simulated Environment

To evaluate the proposed MR-LCMV method, an AV conference room is simulated with the Room Impulse Response (RIR) simulator [12]. The room dimensions are 6m × 4.8m × 3m with reverberation time of 355 milliseconds. A Uniform Linear Array (ULA) consisting of 5 microphones and 0.5m interspacing is considered. Two speakers are sitting symmetrically at a distance of 1.5m from array center at $\theta_1 = 20^o$ and $\theta_2 = 160^o$, for speaker 1 and speaker 2, respectively, as shown in Fig. 3.

For the traditional LCMV, the reference microphone was selected as $r = 1$, that is the microphone closest to speaker 1, marked with a blue circle in Fig. 3. For the MR-LCMV, the reference microphone for speaker 1 remains $r_1 = 1$ and for speaker 2, $r_2 = 5$ is chosen, as marked with blue and red circles in Fig. 3, respectively. The RTFs were estimated using the eigenvalue decomposition (EVD) based method [13].

The simulation was repeated for two types of noise fields: spatially white and diffuse noise fields, with input SNRs of 30 dB and 20 dB respectively, measured on x_1. The diffuse noise was generated using the diffuse noise simulator from [14].

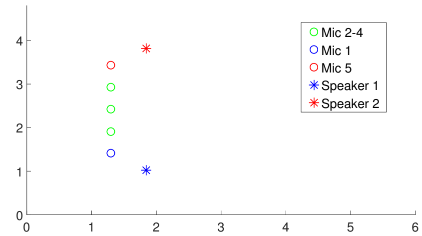


Fig. 3: Simulated conference room.

The STFT parameters was chosen to be $K = 1024$, overlap of 25% and Hamming analysis window. For the norm constraint, we chose $\delta^2 = 15$ dB for both algorithms, and for the MR-LCMV, $\beta^2 = 3$ dB.

The output SNR for each speaker and the normalized distortion energy [11] of speaker i, defined as

$$J_{\text{d}}(\mathbf{w}, i) = |\mathbf{w}^H \mathbf{h}_i - 1|^2, \tag{24}$$

were measured for both scenarios. The results are summarized in Table I. It can be seen that for speaker 2, with the MR-LCMV, we gain approximately 5 dB. This is achieved by maintaining the power of its speech component as received in x_5. The tradeoff here is mainly in the output SNR of speaker 1, which is reduced by approximately 2 dB. This reduction is associated with the amount of noise reduced by both techniques. As to the distortion, for the traditional LCMV $J_{\text{d}}(\mathbf{w}_{\text{LCMV}}, i) = -\infty \; \forall i$, as it solves (17). Some distortion is measured for the MR-LCMV, but it still remains very low.

TABLE I: Simulated environment: Comparison between LCMV and MR-LCMV.

		Traditional		MR-LCMV	
	Noise Type	**Source 1**	**Source 2**	**Source 1**	**Source 2**
SNR [dB]	White	38.2	31.5	36.3	36.3
	Diffuse	28.1	21.3	25.9	26.0
J_{d} [dB]	White	$-\infty$	$-\infty$	-100.2	-105.3
	Diffuse	$-\infty$	$-\infty$	-63.5	-66.7

To visualize the difference between the two methods, we consider the beampowers for both beamformers, as depicted in Fig. 4. The wide-band beampower was created by averaging over all frequencies, as suggested in [15]:

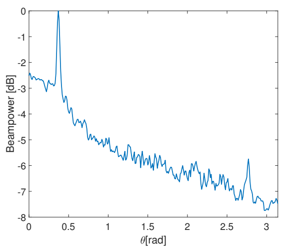
$$\|H(\theta, r = 1.5\text{m})\|^2 = \frac{1}{K} \sum_{k=0}^{K-1} \|\mathbf{w}^H(k) \mathbf{d}(\theta, r = 1.5\text{m}, k)\|^2, \tag{25}$$

where $\mathbf{d}(\theta, r = 1.5\text{m}, k)$ is the generated ATF from angle θ and radius of 1.5 m and $H(\theta, r = 1.5\text{m})$ is the corresponding array response.

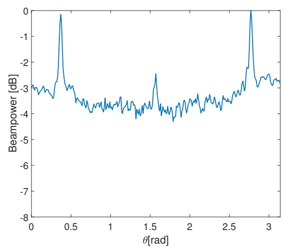
The beampower is generated by (25) for both the LCMV and MR-LCMV and is normalized to have a maximum of 0 dB. Comparing Fig. 4(a) and Fig. 4(b) it can be seen that while the proposed MR-LCMV algorithm maintains equal gain for both desired speakers, the LCMV exhibits a 6 dB attenuation of speaker 2 signal, relatively to speaker 1.

B. Conference Room Recorded Scenario

A distributed microphone array consisting of 2 ULAs distanced at 3m was placed in a large conference room. Each



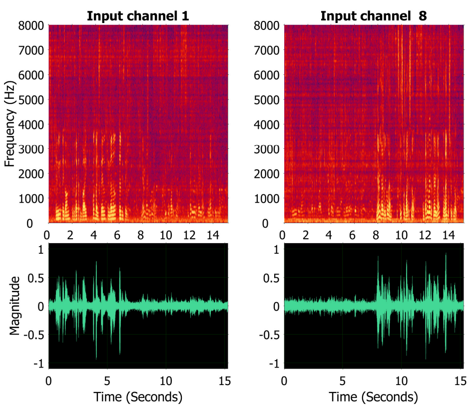
(a)



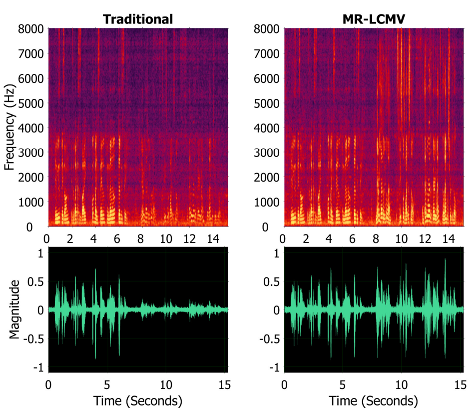
(b)

Fig. 4: Simulated environment: Beampower for (a) traditional LCMV and (b) MR-LCMV algorithms calculated for diffuse noise field.

ULA consists of 4 microphones with 1cm interspacing. A point source interference was placed at 1m from the array center, and an air conditioner was set to high mode. The input signals x_1 and x_8 are depicted in Fig. 5a.



(a)



(b)

Fig. 5: Comparison between the traditional LCMV and MR-LCMV for Conference Room Recorded Scenario: (a) Input for channels x_1 and x_8 and (b) output.

The LCMV and MR-LCMV were designed with the same parameters as in Sec. V-A, with the one change of $r_2 = 8$. The output signals of both the LCMV and the MR-LCMV algorithms are depicted in Fig. 5. Viewing Fig. 5a and comparing it to Fig. 5b, the MR-LCMV property of maintaining the desired speaker energy as captured by the corresponding reference microphones is evident. While for both algorithms speaker 1 level is comparable, the level of speaker 2 produced by the MR-LCMV algorithm is significantly higher. The output SNR values are shown in Table II.

TABLE II: Recorded Scenario: Comparison of output SNR in dB.

Traditional		MR-LCMV	
Source 1	**Source 2**	**Source 1**	**Source 2**
17.5	6.9	15.7	16.8

It can be seen that though, similarly to Sec. V-A, the SNR for speaker 1 is decreased by 2 dB, the SNR for speaker 1 is increased by 10 dB. This is significant improvement for the case of two desired speakers.

REFERENCES

[1] B. Van Veen and K. Buckley, "Beamforming: a versatile approach to spatial filtering," *ASSP Magazine, IEEE*, vol. 5, no. 2, pp. 4–24, Apr. 1988.

[2] H. L. Van Trees, *Optimum array processing: Part IV of detection, estimation, and modulation theory*. John Wiley & Sons, 2004.

[3] A. Bertrand and M. Moonen, "Distributed node-specific LCMV beamforming in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 233–246, 2011.

[4] 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, Aug. 2001.

[5] K. Reindl, S. Markovich-Golan, H. Barfuss, S. Gannot, and W. Kellermann, "Geometrically constrained TRINICON-based relative transfer function estimation in underdetermined scenarios," in *2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. IEEE, 2013, pp. 1–4.

[6] S. Markovich, S. Gannot, and I. Cohen, "Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signals," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 6, pp. 1071–1086, 2009.

[7] S. Gannot, E. Vincent, S. Markovich-Golan, and A. Ozerov, "A consolidated perspective on multimicrophone speech enhancement and source separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 4, pp. 692–730, 2017.

[8] T. C. Lawin-Ore and S. Doclo, "Reference microphone selection for MWF-based noise reduction using distributed microphone arrays," in *Speech Communication; 10. ITG Symposium*. VDE, 2012, pp. 1–4.

[9] H. Cox, R. Zeskind, and M. Owen, "Robust adaptive beamforming," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 10, pp. 1365–1376, 1987.

[10] A. Barnov, V. B. Bracha, S. Markovich-Golan, and S. Gannot, "Spatially robust GSC beamforming with controlled white noise gain," in *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*. IEEE, 2018, pp. 231–235.

[11] S. Doclo and M. Moonen, "Superdirective beamforming robust against microphone mismatch," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 2, pp. 617–631, 2007.

[12] E. A. Habets, "Room impulse response generator," *Technische Universiteit Eindhoven, Tech. Rep*, vol. 2, no. 2.4, p. 1, 2006.

[13] R. Varzandeh, M. Taseska, and E. A. Habets, "An iterative multichannel subspace-based covariance subtraction method for relative transfer function estimation," in *2017 Hands-free Speech Communications and Microphone Arrays (HSCMA)*. IEEE, 2017, pp. 11–15.

[14] E. A. Habets, I. Cohen, and S. Gannot, "Generating nonstationary multisensor signals under a spatial coherence constraint," *The Journal of the Acoustical Society of America*, vol. 124, no. 5, pp. 2911–2917, 2008.

[15] J. Dmochowski, J. Benesty, and S. Affès, "On spatial aliasing in microphone arrays," *IEEE Transactions on Signal Processing*, vol. 57, no. 4, pp. 1383–1395, 2008.