LCMV BEAMFORMING WITH SUBSPACE PROJECTION FOR MULTI-SPEAKER SPEECH ENHANCEMENT

Amin Hassani, Alexander Bertrand, Marc Moonen

KU Leuven, Dept. of Electrical Engineering-ESAT,
Stadius Center for Dynamical Systems, Signal Processing and Data Analytics,
Address: Kasteelpark Arenberg 10, B-3001 Leuven, Belgium
E-mail: amin.hassani, alexander.bertrand, marc.moonen@esat.kuleuven.be

ABSTRACT

The linearly constrained minimum variance (LCMV) beamformer has been widely employed to extract (a mixture of) multiple desired speech signals from a collection of microphone signals, which are also polluted by other interfering speech signals and noise components. In many practical applications, the LCMV beamformer requires that the subspace corresponding to the desired and interferer signals is either known, or estimated by means of a data-driven procedure, e.g., using a generalized eigenvalue decomposition (GEVD). In practice, however, it often occurs that insufficient relevant samples are available to accurately estimate these subspaces, leading to a beamformer with poor output performance. In this paper we propose a subspace projection-based approach to improve the performance of the LCMV beamformer by exploiting the available data more efficiently. The improved performance achieved by this approach is demonstrated by means of simulation results.

Index Terms— LCMV beamforming, generalized eigenvalue decomposition, subspace estimation, speech enhancement, noise reduction.

1. INTRODUCTION

Sensor arrays allow space-time signal processing which often improves the performance of parameter- or signal-of-interest estimation, when compared with single-sensor based estimation [1, 2]. In audio and speech enhancement applications, microphone arrays have been widely used [3]. A common problem is to extract (a mixture of) multiple desired speech signals from the microphone signals, which are also polluted by other interfering speech signals and noise components. To solve this problem, one can use a so-called beamforming approach [4]. A well-known approach is linearly constrained minimum variance (LCMV) beamforming which aims at minimizing the total power of the beamformer output, under a set of linear constraints that control the array beam pattern such that the signals coming from the desired directions remain undistorted while signals coming from the interfering directions are rejected [4].

Basically, there are two main classes of LCMV beamformers. The first class assumes that each individual room impulse response (RIR) (or equivalently acoustic transfer function (ATF)) between each source and each microphone is known [5]. In this case, the LCMV beamformer will estimate the mixture of desired source signals (that have not yet been distorted by the RIRs). The second class deals with cases where the RIRs are not known *a priori* and hence have to be estimated on the fly based on statistical properties of the microphone signals. This class is often referred to as *blind* LCMV beamforming [5, 6]. In practice however, estimating individual RIRs may not be straightforward, as it is usually required that there are sufficient signal segments in which only *one* of the sources is active, i.e., for each of the individual desired and interfering sources [7]. In [5, 6], the authors proposed a beamforming framework in which the unknown ATFs of the desired and interfering sources are replaced by respective bases for the desired sources and interfering sources subspaces spanned by the columns of the true ATFs. The resulting response then estimates the mixture of the desired source signals as observed by an arbitrarily chosen *reference microphone* and suppresses the interfering source signals. In this paper we consider such an LCMV beamformer for which the desired sources and interfering sources subspaces must be estimated based on the microphone signals.

To estimate the desired sources and interfering sources subspaces, an eigenvalue decomposition- (EVD-) based ap-

The authors would like to thank Prof. S. Gannot for the interesting discussions on the topic of this paper.

This work was carried out at the ESAT Laboratory of KU Leuven, in the frame of KU Leuven Research Council CoE PFV/10/002 (OPTEC), BOF/STG-14-005, the Interuniversity Attractive Poles Programme initiated by the Belgian Science Policy Office IUAP P7/23 (BESTCOM), Research Project FWO nr. G.0763.12 'Wireless Acoustic Sensor Networks for Extended Auditory Communication', 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', and project 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.

proach can be used (as in [8]). However, a generalized EVD (GEVD-) based subspace estimation is better suited for scenarios with spatially correlated noise, as it directly incorporates the estimated noise covariance matrix such that each resulting subspace estimate is aimed to have the highest output signal to noise ratio (SNR) [9, 10].

For the estimation of the desired sources and interfering sources subspaces, in practice we must first construct the relevant sample covariance matrices based on the time segments during which only the desired or interfering sources are active, namely *'desired-sources-only'* and *'interfering-sources-only'* segments, respectively. This procedure in practice requires a voice activity detector (VAD) that is able to distinguish between such segments (e.g., as in [11]). Note that in this way the samples from the segments during which both the desired source(s) and the interfering source(s) are simultaneously active will be discarded for the estimation of the individual subspaces. In practice, however, it often happens that insufficient 'desired-sources-only' and 'interfering-sources-only' samples are available to accurately estimate these individual subspaces. In this paper we propose a subspace projection-based approach which improves the output performance of the blind LCMV beamformer based on the projection of the individual subspace estimates onto the joint signal subspace of all the desired and interfering sources present in the environment. Basically, the motivation behind this is the fact that now all segments can be involved for the estimation of the joint signal subspace (except for '*noise-only*' segments). Hence the accuracy and tracking performance of this joint subspace estimation is expected to be higher compared to the individual subspace estimates.

2. DATA MODEL AND PROBLEM STATEMENT

We consider a microphone array with M microphones, in which the captured signal at microphone m, $m = 1, \ldots, M$ can be described in the frequency domain as

$$y_m(\omega) = d_m(\omega) + i_m(\omega) + n_m(\omega) \tag{1}$$

where $d_m(\omega)$ is the desired source signals component and $i_m(\omega)$ is the interfering speech signals component, and where $n_m(\omega)$ denotes the additive noise component which includes both spatially correlated and uncorrelated noise contributions. Although $i_m(\omega)$ can also be considered as noise, it is not included in $n_m(\omega)$, because we aim to control the suppression of the interferers, possibly targeting a complete removal. In (1), ω is the discrete frequency-domain variable where the resolution is defined by the discrete Fourier transform (DFT) of size L. For the sake of brevity, ω will be omitted in the sequel. We assume that there are N_d desired speech sources, and N_i interfering speech sources and that these numbers are known (although they could also be estimated in practice). Hence $d_m = \sum_{d=1}^{N_d} a_{dm} s_d$ and $i_m = \sum_{i=1}^{N_i} a_{im} s_i$, where a_{dm} and a_{im} denote the ATFs from the desired speech source s_d and the interfering speech source s_i to microphone m, respectively. The stacked version of all microphone signals is represented as

$$\mathbf{y} = \mathbf{A}_d \mathbf{s}_d + \mathbf{A}_i \mathbf{s}_i + \mathbf{n} \triangleq \mathbf{d} + \mathbf{i} + \mathbf{n} \tag{2}$$

where $\mathbf{A}_d = [\mathbf{a}_{d_1} \ldots \mathbf{a}_{d_{N_d}}]$, $\mathbf{A}_i = [\mathbf{a}_{i_1} \ldots \mathbf{a}_{i_{N_i}}]$ are $M \times N_d$ and $M \times N_i$ steering matrices, respectively, with $\mathbf{a}_x$ denoting the RIR (ATF) from the source x to the microphone array. In (2), $\mathbf{s}_d$ and $\mathbf{s}_i$ are stacked signal vectors containing the N_d desired speech source signals and N_i interfering speech source signals, respectively.

In this paper we consider the problem of extracting the mixture of the desired speech signals as it is observed at the reference microphone, from the noisy microphone signals $\mathbf{y}$ and with an LCMV beamformer. This extraction is assumed to be carried out in scenarios where insufficient 'desired-sources-only' and 'interfering-sources-only' samples are available to accurately estimate the individual subspaces spanned by the columns of $\mathbf{A}_d$ and $\mathbf{A}_i$, respectively.

3. LCMV BEAMFORMING

LCMV beamforming in general applies a linear M-dimensional estimator $\mathbf{w}$ to the M-channel signal $\mathbf{y}$ to estimate the desired signal $\bar{d} = \mathbf{w}^H \mathbf{y}$, where H denotes the conjugate transpose operator, and where overline $\bar{(.)}$ denotes the estimate. To design an LCMV beamformer that estimates the unreverberated source signals $\mathbf{s}_d$, the steering matrices $\mathbf{A}_d$ and $\mathbf{A}_i$ have to be known [5]. When instead of estimating $\mathbf{s}_d$ the aim is to estimate mixture of the desired speech signals as captured by the reference microphone, a modified LCMV beamformer can be designed which requires only estimates of $\mathbb{Q}_d$ and $\mathbb{Q}_i$, where $\mathbb{Q}_d$ is an $M \times N_d$ matrix at which the columns define a unitary basis for the desired sources subspace spanned by the columns of $\mathbf{A}_d$, and where $\mathbb{Q}_i$ is an $M \times N_i$ matrix where its columns define a unitary basis for the interfering sources subspace spanned by the columns of $\mathbf{A}_i$ [5, 6]. The resulting LCMV problem can be expressed as [6]

$$\min_{\mathbf{w}} \; E\{|\mathbf{w}^H \mathbf{y}|^2\} \tag{3}$$

$$s.t. \; \mathbb{Q}^H \mathbf{w} = \mathbf{f} \tag{4}$$

where $\mathbb{Q} \triangleq [\mathbb{Q}_d \; \mathbb{Q}_i]$, and where $\mathbf{f}$ is the vector of desired responses defined as $\mathbf{f} = [\mathbb{q}_d^T \; \mathbf{0}]^T$ where $\mathbb{q}_d$ is the j-th column of $\mathbb{Q}_d^H$, with j denoting the reference microphone. In the sequel and without loss of generality (w.l.o.g.), we assume that the first microphone is chosen as the reference microphone, i.e., $j = 1$. The solution of (3)-(4) is then given by

$$\mathbf{w} = \mathbf{R}_{yy}^{-1} \mathbb{Q} (\mathbb{Q}^H \mathbf{R}_{yy}^{-1} \mathbb{Q})^{-1} \mathbf{f}. \tag{5}$$

Note that (5) has to be computed for each frequency bin separately. The resulting output signal, namely $\bar{d}_{\text{ref}}$, can be then described as $\bar{d}_{\text{ref}} = \mathbf{w}^H \mathbf{y} = \sum_{d=1}^{N_d} a_{d1} s_d + \hat{\mathbf{w}}^H \mathbf{n}$, which verifies the fact that this solution estimates the mixture of the desired speech signals as captured by the first microphone, while fully cancelling the interfering speech signals and while suppressing the ambient noise as much as possible [6].

To estimate $\mathbb{Q}_d$ and $\mathbb{Q}_i$ based on the microphone signal $\mathbf{y}$ in (2), we first define the following source-activity-based correlation matrices:

$$\mathbf{R}_{yy}^d = \mathbf{A}_d \mathbf{\Pi}_d \mathbf{A}_d^H + \mathbf{R}_{nn} \tag{6}$$

$$\mathbf{R}_{yy}^i = \mathbf{A}_i \mathbf{\Pi}_i \mathbf{A}_i^H + \mathbf{R}_{nn} \tag{7}$$

where $\mathbf{\Pi}_d = \text{diag}\{P_{d_1} \ldots P_{d_{N_d}}\}$ and $\mathbf{\Pi}_i = \text{diag}\{P_{i_1} \ldots P_{i_{N_i}}\}$, with P_x being the power of x-th source signal, and where $\mathbf{R}_{nn} = E\{\mathbf{n}\mathbf{n}^H\}$. Note that in practice the correlation matrices $\mathbf{R}_{yy}^d$ and $\mathbf{R}_{yy}^i$ can be estimated via sample averaging over the 'desired-sources-only' and 'interfering-sources-only' segments, respectively, requiring an oracle algorithm that can distinguish between these segments [5, 6]. An EVD-based approach can then be used to estimate the subspaces (e.g., as in [5,6,8]). In [5,6], the authors proposed a procedure to choose a set of N_d and N_i eigenvectors (EVCs) of $\mathbf{R}_{yy}^d$ and $\mathbf{R}_{yy}^i$ that span the same subspace as $\mathbb{Q}_d$ and $\mathbb{Q}_i$, respectively.

4. PROJECTION-BASED SUBSPACE ESTIMATION

The estimation of $\mathbb{Q}_d$ and $\mathbb{Q}_i$, as explained in Section 3, may yield poor results if (6) and (7) can not be accurately estimated, e.g., when there are insufficient 'desired-sources-only' and/or 'interfering-sources-only' segments or samples available. Indeed, in the procedure described in Section 3, a large part of the data is not used, namely the signal segments in which desired and interfering sources are simultaneously active. In this section, we propose a method which also exploits these signal segments, which leads to an improved speech enhancement performance.

Here we employ a GEVD-based subspace estimation although a similar strategy can be used for other subspace estimation techniques. Define $\mathbf{X}_d$ and $\mathbf{X}_i$ as $M \times M$ matrices containing the generalized EVCs (GEVCs) of the ordered matrix pair $(\mathbf{R}_{yy}^d, \mathbf{R}_{nn})$ and $(\mathbf{R}_{yy}^i, \mathbf{R}_{nn})$, respectively, in their columns. Note that $\mathbf{R}_{nn}$ can be estimated from the 'noise-only' segments when all the desired and interfering speech sources are inactive. We assume (w.l.o.g.) that 1) the GEVCs are sorted such that their corresponding generalized eigenvalues (GEVLs) are sorted in descending order 2) the GEVCs are scaled such that $\mathbf{X}^H \mathbf{R}_{nn} \mathbf{X} = \mathbf{I}_M$. Now let $\mathbf{Q}_d = (\mathbf{X}_d)^{-H}$ and $\mathbf{Q}_i = (\mathbf{X}_i)^{-H}$. It can then be verified that the first N_d columns of $\mathbf{Q}_d$ and the first N_i columns of $\mathbf{Q}_i$ span the same subspace as $\mathbb{Q}_d$ and $\mathbb{Q}_i$, respectively [9].

As mentioned earlier, because of insufficient 'desired-sources-only' or 'interfering-sources-only' segments, $\mathbb{Q}_d$ and $\mathbb{Q}_i$ will be poorly estimated, which may often result in inadequate LCMV beamforming outputs. In such conditions we propose the following subspace projection-based approach such that the discarded samples associated with the segments during which the desired and interfering sources are simultaneously active can also be exploited. Excluding the samples of the 'noise-only' segments, all other segments are then used to estimate

$$\mathbf{R}_{yy}^{d,i} = \mathbf{A}_d \mathbf{\Pi}_d \mathbf{A}_d^H + \mathbf{A}_i \mathbf{\Pi}_i \mathbf{A}_i^H + \mathbf{R}_{nn}. \tag{8}$$

We now define $\mathbf{X}_{d,i}$ as the full-rank matrix containing the GEVCs of the ordered matrix pair $(\mathbf{R}_{yy}^{d,i}, \mathbf{R}_{nn})$. The joint $(N_d + N_i)$-dimensional desired sources and interfering sources subspace $\mathbb{Q}_{d,i}$ can then be defined as the first $(N_d + N_i)$ columns of the matrix $\mathbf{Q}_{d,i} = (\mathbf{X}_{d,i})^{-H}$.

Note that in theory, the columns of $\mathbb{Q}_{d,i}$ and the columns of $[\mathbb{Q}_d \ \mathbb{Q}_i]$ span the same signal subspace. In practice however, because of the discrepancies between the data segments based on which the correlation matrices (6)-(8) are estimated, this does not hold anymore. This can be corrected by the projection of the poorly estimated $\mathbb{Q}_d$ and $\mathbb{Q}_i$ onto the joint subspace estimate $\mathbb{Q}_{d,i}$. Hence we define the projected individual subspace estimates as

$$\mathbb{Q}_d^{\text{proj}} \triangleq \mathbb{Q}_{d,i}(\mathbb{Q}_{d,i}^T \mathbb{Q}_{d,i})^{-1} \mathbb{Q}_{d,i}^T \mathbb{Q}_d \tag{9}$$

$$\mathbb{Q}_i^{\text{proj}} \triangleq \mathbb{Q}_{d,i}(\mathbb{Q}_{d,i}^T \mathbb{Q}_{d,i})^{-1} \mathbb{Q}_{d,i}^T \mathbb{Q}_i \tag{10}$$

The subspace projection-based version of the LCMV beamformer solution (5) can then be expressed as

$$\mathbf{w}_{\text{proj}} = (\mathbf{R}_{yy}^{d,i})^{-1} \mathbb{Q}_{\text{proj}} (\mathbb{Q}_{\text{proj}}^H (\mathbf{R}_{yy}^{d,i})^{-1} \mathbb{Q}_{\text{proj}})^{-1} \mathbf{f}_{\text{proj}} \tag{11}$$

where $\mathbb{Q}_{\text{proj}} \triangleq [\mathbb{Q}_d^{\text{proj}} \ \mathbb{Q}_i^{\text{proj}}]$ and where $\mathbf{f}_{\text{proj}} \triangleq [\mathbb{q}_{\text{proj}}^T \ \mathbf{0}]^T$, with $\mathbb{q}_{\text{proj}}$ being the first column of $(\mathbb{Q}_d^{\text{proj}})^H$. The actual output of the beamformer (11), i.e., $\bar{d}_{\text{proj}} = \mathbf{w}_{\text{proj}}^H \mathbf{y}$, will be evaluated in the next section via simulation results.

5. SIMULATION RESULTS

In this section, the improved performance achieved by the subspace projection-based LCMV solution (11) is demonstrated by means of simulation results. For this goal, two different scenarios are simulated. The first scenario assumes multiple desired and multiple interfering sources in the enclosure, with narrowband source signals. This scenario allows us to easily perform Monte Carlo (MC) simulations to better investigate the benefits of the proposed approach in different conditions. The second scenario tests the proposed approach for multi-talker speech enhancement where the desired and interfering sources produce different speech signals (English sentences).

5.1. Simulated scenario with narrowband source signals

A setup with different position of nodes and sources, and with different narrowband source signals is considered in each MC run. Further specifications of this scenario are as follows: $M = 10$, $N_d = 2$, $N_i = 3$, total number of samples$= 20000$, number of samples in which both desired and interfering sources are active$= 7000$ and MC runs$= 1000$. Number of available 'desired-sources-only' and 'interfering-sources-only' samples, namely Nb_{only}, are assumed to be equal. Nb_{only} is then varied from 1 to 5000 (see Figure 1). The remaining $13000 - 2Nb_{\text{only}}$ samples are 'noise-only'. All desired and interfering sources have the same power P. The noise consists of two randomly placed spatial noise sources with power $0.5P$, as well as uncorrelated noise on each sensor which is 5% of the power of the first desired source as

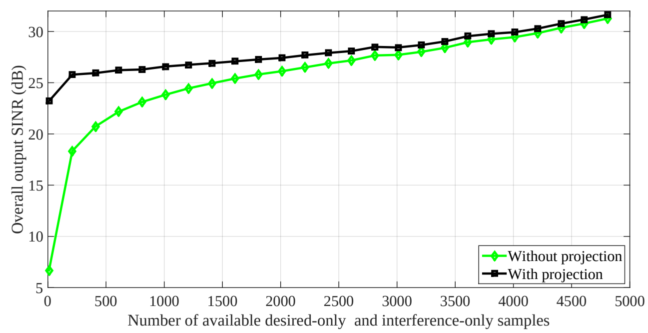

Fig. 1. MC results based on narrowband source signals

observed on the first sensor. The entries of the steering matrices are independently drawn from a uniform distribution over the interval $[0.5; 0.5]$. The same holds for the samples of all the involved source signals, followed by a proper scaling to modify their power. As a performance measure, we utilized the output *signal to interference plus noise ratio* (oSINR) at the reference sensor, defined as

$$\text{oSNRI} = 10 \log_{10} \frac{E\{|\mathbf{w}^H \mathbf{d}|^2\}}{E\{|\mathbf{w}^H \mathbf{i}|^2\} + E\{|\mathbf{w}^H \mathbf{n}|^2\}} \tag{12}$$

(expectations are taken over all frequency-time points). Figure 1 compares the output oSNRI of the proposed subspace projection-based LCMV beamformer (11) to that of (5), as a function of the number of available 'desired-sources-only' and 'interference-sources-only' samples. As can be seen, the proposed approach significantly outperforms when insufficient 'desired-sources-only' and 'interference-only' samples are available. Note that two figures eventually converge to each other when sufficiently large number of relevant samples are available.

5.2. Multi-talker speech enhancement

In this scenario we simulate a cubic room with dimensions $5m \times 5m \times 5m$ and with surface reflection coefficient $\beta = 0.2$ using the image method [12]. The RIRs were simulated based on the modified version in [13]. A uniform linear microphone array consisting of $M = 10$ omni-directional microphones with inter-microphone distance of $5cm$ is considered where the center microphone is located at the position $[x = 2.5m, y = 1.5m]$. A desired speech source, an interfering speech source and a babble noise source is located at $[x = 1m, y = 2m]$, $[x = 4m, y = 2m]$ and $[x = 2.5m, y = 3.5m]$, respectively. We use a sampling frequency of $F_s = 16kHz$, a Hann-windowed DFT with size $L = 512$ and with 50% overlap. To avoid including the effect of VAD errors, we here use an ideal VAD with the ability of distinguishing between the desired and interfering speech sources. Both the desired and interfering speech sources produce short sentences with the same power $P_s = P_i$, with 7 seconds of overlapping activity and with some silence periods in between sentences (see top plot of Figure 2). The power of the babble noise source is $0.5P_s$. An additional spatially uncorrelated noise component at each microphone is simulated with a white Gaussian

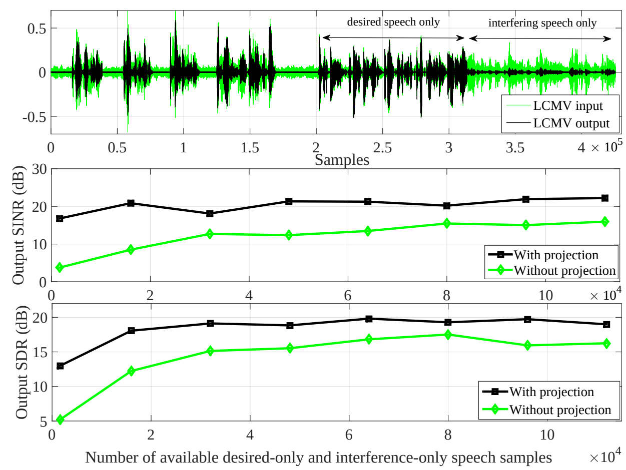

Fig. 2. Results based on a simulated room with speech signals

signal with 5% of the power of the desired speech signal as observed at the first microphone. To evaluate the performance, we again increase the number of available samples in 'desired-sources-only' and 'interfering-sources-only' segments, varying from $0.1F_s$ to $7F_s$. Besides oSINR, we here also consider the output signal to distortion ratio (oSDR) at the first microphone, defined as

$$\text{oSDR} = 10 \log_{10} \frac{E\{|\mathbf{d}|^2\}}{E\{|\mathbf{d} - \mathbf{w}^H \mathbf{d}|^2\}} \tag{13}$$

In the simulated scenario, input SNR$\approx 9.5dB$, input SIR$\approx 2.2dB$ and input SINR$\approx 1.5dB$, measured at the first microphone. The middle and bottom part of Figure 2 evaluate the performance of the LCMV beamformer output with the projection-based approach in terms of the output SINR and SDR. These convincing results again verify that the proposed projection-based approach delivers a significantly better performance. This improvement is indeed achieved at the cost of more complex computations due to the need for the computation of the full joint subspace $\mathbb{Q}_{d,i}$ which in turn requires to perform an extra GEVD. Note that a sufficiently large number of available samples lets the plots in Figure 2 to converge to each other (not shown here).

6. CONCLUSION

In this paper, we have proposed a subspace projection-based approach to increase the performance of an LCMV beamformer in conditions where insufficient relevant samples are available to accurately estimate the subspaces of the desired sources and interfering sources, respectively. We have considered a GEVD-based method for subspace estimation in combination with a subspace projection step, which allows to better estimate the desired sources and interfering sources subspaces. This improvement is achieved at the cost of more complex computations, as the poorly estimated subspaces have to be projected onto the larger joint subspace, which itself requires an extra GEVD. The improved performance achieved by this subspace projection-based approach has been demonstrated by means of simulation results.

7. REFERENCES

[1] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *Signal Processing Magazine, IEEE*, vol. 13, no. 4, pp. 67–94, 1996.

[2] S.U. Pillai and C.S. Burrus, *Array signal processing*, Signal Processing and Digital Filtering. Springer-Verlag, 1989.

[3] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*, Digital Signal Processing - Springer-Verlag. Springer, 2001.

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

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

[6] S.M. Golan, S. Gannot, and I. Cohen, "Subspace tracking of multiple sources and its application to speakers extraction," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, March 2010, pp. 201–204.

[7] J. Benesty, Jingdong Chen, Yiteng Huang, and J. Dmochowski, "On microphone-array beamforming from a MIMO acoustic signal processing perspective," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 15, no. 3, pp. 1053–1065, March 2007.

[8] Y. Ephraim and H.L. Van Trees, "A signal subspace approach for speech enhancement," *Speech and Audio Processing, IEEE Transactions on*, vol. 3, no. 4, pp. 251–266, Jul 1995.

[9] A. Hassani, A. Bertrand, and M. Moonen, "Distributed GEVD-based signal subspace estimation in a fully-connected wireless sensor network," in *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*, Sept 2014, pp. 1292–1296.

[10] 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," *Audio, Speech, and Language Processing, IEEE/ACM Transactions on*, vol. 22, no. 4, pp. 785–799, April 2014.

[11] A. Bertrand and M. Moonen, "Energy-based multi-speaker voice activity detection with an ad hoc microphone array," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, March 2010, pp. 85–88.

[12] J. Allen and D. Berkley, "Image method for efficiently simulating smallroom acoustics," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.

[13] E. Habets, "Room impulse response (RIR) generator," 2010.