ℓ_1-CONSTRAINED MVDR-BASED SELECTION OF NONIDENTICAL DIRECTIVITIES IN MICROPHONE ARRAY

Satoru Emura

NTT Media Intelligence Laboratories, NTT Corporation

ABSTRACT

We propose a method for finding a good combination of nonidentical directivities for a microphone array, i.e., a combination that can achieve better noise suppression. To avoid checking all combinations of directivities, the method incrementally applies ℓ_1-constrained minimum variance distortionless response beamforming to the coherently-focused power spectral density matrices of microphone array signals. The ℓ_1 constraint selects microphones with greater importance by inducing a sparse filter for beamforming.

Index Terms— microphone array, nonidentical, MVDR, ℓ_1 constraint

1. INTRODUCTION

Hands-free or distant audio acquisition is indispensable for many applications such as teleconferencing and audio recording. The signals captured using multiple microphones are processed in order to extract the desired sound by beamforming, which suppresses interference and noise [1].

The minimum variance distortionless response (MVDR) beamformer and linearly constrained minimum variance (LCMV) beamformer for audio signal processing have been intensively studied over the past decade [2][3][4]. The MVDR beamformer minimizes its output power under the constraint of no signal distortion. In the LCMV beamformer, additional linear constraints are imposed.

It is quite common in the literature to use an array consisting of identical omnidirectional sensors or microphones. Recently, combining sensors with different directivities has begun to be studied. Levin et al. [5] discussed robust beamforming of an array of sensors with nonidentical directivity patterns. Cao et al. [6] discussed the MVDR beamformer for an acoustic vector sensor that consists of one monopole and three dipole sensors.

For obtaining a better sound acquisition, both an appropriate directivity combination and an appropriate filter for beamforming are required, where to efficiently find a good directivity combination, i.e., a combination that achieves better noise reduction, is an issue because: a) the number of combinations drastically increases as the number of microphones increases, b) checking all the combinations to find better combinations is difficult with an array with many microphones, and c) once a good combination is obtained, a good filter for beamforming can be derived by MVDR.

We propose a method for efficiently finding a good combination without checking all the combinations of microphone directivities. This method uses ℓ_1-constrained MVDR, which can induce a sparse solution of a filter for beamforming. The method starts from the situation in which M microphones of directivity A and M microphones of directivity B are collocated redundantly for M microphone positions. The microphones with greater importance are incrementally selected. The computational complexity is further reduced by applying this procedure to the power spectral density (PSD) matrices after focusing [7] .

2. ARRAY MODEL

Directivity model

First, we assume that the directivity pattrens of each microphone element can be modeled as a differential microphone. Standard directivity patterns, such as cardioid, hypercardioid and supercardioid, can be modeled as the first-order differential microphones, which delay and subtract the signals of two closely spaced omnidirectional microphones [8].

Hence, without loss of generality, we consider an array of $2M$-element omnidirectional microphones instead of an array with M microphones of directivity A and M microphones of directivity B.

Signal model

Consider the conventional signal model with which an $2M$-omnidirectional-element microphone array captures convolved target signals in a noise field. We assume that the mth microphone ($m = 1, \dots M$) is closely spaced with the $(m + M)$th microphone. The captured signals are expressed as

$$\begin{aligned} y_m(k) &= g_m * s(k) + v_m(k) \\ &= x_m(k) + v_m(k), \quad m = 1, 2, \dots, 2M, \end{aligned} \tag{1}$$

where g_m is the impulse response from the unknown source $s(k)$ to the mth microphone, $*$ stands for convolution, and $v_m(k)$ is the noise at mth microphone. We assume that all signals considered are broadband and that the signals $x_m(k)$

and $v_m(k)$ are uncorrelated and zero mean. In the frequency domain, (1) can be rewritten as

$$\begin{aligned} Y_m(\omega) &= G_m(\omega)S(\omega) + V_m(\omega) \\ &= X_m(\omega) + V_m(\omega), \quad m = 1, 2, \ldots, 2M, \end{aligned} \tag{2}$$

where $Y_m(\omega)$, $G_m(\omega)$, $S(\omega)$, $X_m(\omega)$ and $V_m(\omega)$ are the short-term Fourier transforms (STFTs) of $y_m(k)$, g_m, $s(k)$, $x_n(k)$, and $v_n(k)$ at frequency ω. In a vector notation, (2) can be rewritten as

$$\mathbf{y}(\omega) = \mathbf{x}(\omega) + \mathbf{v}(\omega), \tag{3}$$

where $\mathbf{y}(\omega)$, $\mathbf{x}(\omega)$, and $\mathbf{v}(\omega)$ are the column vectors of $Y_m(\omega)$, $X_m(\omega)$, $V_m(\omega)$ $(m = 1, 2, \ldots, 2M)$. The narrow-band PSD matrices of noise and target at frequency ω are given as

$$\mathbf{R}_V(\omega) = E\left[\mathbf{v}(\omega)\mathbf{v}^H(\omega)\right], \tag{4}$$

$$\mathbf{R}_X(\omega) = E\left[\mathbf{x}(\omega)\mathbf{x}^H(\omega)\right], \tag{5}$$

where superscript ^{H} denotes the transpose conjugation of a vector or a matrix.

Focusing

Focusing is a technique originally proposed for coherent signal-subspace processing for direction-of-arrival estimation of wide-band sources, where the signal subspaces of the narrow-band PSD matrices are aligned and averaged into a single PSD [9] . We use the focusing matrices developed by Doron et al. [10] and Doron and Doron [11]. Let $\mathbf{T}(\omega_i)$ be the focusing matrix from frequency ω_i to the focusing frequency ω_f (see 3.4 for the detail). The focused PSD matrices are given as follows.

$$\mathbf{R}_V(\omega_f) = \sum_i \mathbf{T}(\omega_i)\mathbf{R}_V(\omega_i)\mathbf{T}^H(\omega_i) \tag{6}$$

$$\mathbf{R}_X(\omega_f) = \sum_i \mathbf{T}(\omega_i)\mathbf{R}_X(\omega_i)\mathbf{T}^H(\omega_i) \tag{7}$$

After focusing, the directivity of each microphone is restored using the differential microphone model. Its output is

$$X_m^a(\omega) = \frac{\left(X_m(\omega) - X_{m+M}(\omega)e^{-j\omega\tau_a}\right)}{1 - e^{-j\omega(\tau_a+\tau_d)}}, \tag{8}$$

where τ_d is $d/(sound\,velocity)$, τ_a determines directivity A, d is the distance between closely spaced omnidirectional microphones, and j is the imaginary unit. The bidirectional, cardioid, and hypercardioid directivity patterns can be obtained by $\tau_a = 0$, $\tau_a = \tau_d$, $\tau_a = \tau_d/3$ respectively. Let $\mathbf{D}(\omega)$ be the matrix that relates identical omnidirectional microphone signals to nonidentical directivity microphone signals based on the differential microphone model. The PSD matrices of a nonidentical directivity microphone array are expressed as

$$\mathbf{R}_V^{ab}(\omega_f) = \mathbf{D}(\omega_f)\mathbf{R}_V(\omega_f)\mathbf{D}(\omega_f)^H \tag{9}$$

$$\mathbf{R}_X^{ab}(\omega_f) = \mathbf{D}(\omega_f)\mathbf{R}_\mathbf{X}(\omega_f)\mathbf{D}(\omega_f)^H. \tag{10}$$

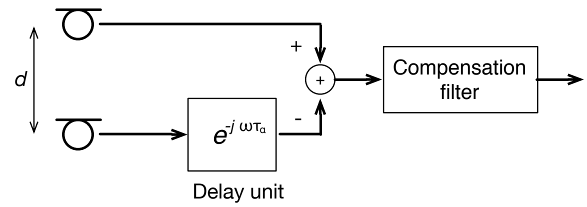

Fig. 1. First-order differential microphone model

3. MICROPHONE SELECTION

We propose a method for efficiently finding a good combination without checking all the combinations of microphone directivities. Microphones with greater importance are selected by incrementally applying ℓ_1-constrained MVDR beamforming to the focused PSD matrices of (9) and (10).

Focusing is essential because wide-band signals is necessary to determine the combination of microphone directivities and multiple narrow-band signals are difficult to handle simultaneously. Focusing can drastically reduce the amount of narrow-band signal data and make the selection procedure efficient.

We use Doron's focusing technique [10]. Since this focusing technique requires that all the microphones are identically omnidirectional, we apply the differential microphone model for the transformation between identical and nonidentical microphone arrays.

In this section, we first show how we add the ℓ_1 constraint to MVDR beamforming. We then explain the procedure of directivity selection and details of focusing.

3.1. Conventional MVDR

Let $\mathbf{h}(\omega)$ be the beamforming weight vector. The beamformer output $Z(\omega)$ is obtained by applying a complex weight to each microphone signal and summing

$$Z(\omega) = \mathbf{h}^H(\omega)\left\{\mathbf{x}^{ab}(\omega) + \mathbf{v}^{ab}(\omega)\right\}, \tag{11}$$

where $\mathbf{x}^{ab}(\omega)$ is the vector consisting of the directivity-A microphone signals $X_1^a(\omega), \cdots, X_M^a(\omega)$ and the directivity-B microphone signals $X_2^b(\omega), \cdots, X_M^b(\omega)$ of the target source, and $\mathbf{v}^{ab}(\omega)$ is that of noise. We consider the first microphone with directivity A as the reference microphone. The conventional MVDR beamformer is then given as

$$\mathbf{h}_{MVDR}(\omega) = arg\min \mathbf{h}^H(\omega)\mathbf{R}_V^{ab}(\omega)\mathbf{h}(\omega) \tag{12}$$

$$\text{subject to } \mathbf{h}^H(\omega)\mathbf{g}^{ab}(\omega) = G_1^a(\omega), \tag{13}$$

where $\mathbf{g}^{ab}(\omega)$ is the vector of transfer functions of the acoustic path from the target source to each microphone and $G_1^a(\omega)$ is that to the reference microphone.

By multiplying $s(\omega)s^*(\omega)G_1^{a*}(\omega)$ with (13) from the right and taking expectation by $E[\bullet]$, we can rewrite the constraint as

$$\mathbf{h}^H(\omega)E\left[\mathbf{x}^{\mathbf{ab}}(\omega)X_1^{a*}(\omega)\right] = E\left[X_1^a(\omega)X_1^{a*}(\omega)\right], \quad (14)$$

where $E[\mathbf{x}^{\mathbf{ab}}(\omega)X_1^{a*}(\omega)]$ is given as (:,1) elements of the matrix $\mathbf{R}_X^{ab}(\omega)$ and $E[X_1^a(\omega)X_1^{a*}(\omega)]$ is given as its (1,1) element in the Matlab style.

3.2. ℓ_1-constrained MVDR

We add the ℓ_1 constraint as an inequality constraint (17) to MVDR beamforming according to Hastie [12].

$$\mathbf{h}_\rho(\omega) = arg\min \mathbf{h}^H(\omega)\mathbf{R}_V^{ab}(\omega)\mathbf{h}(\omega) \quad (15)$$

subject to

$$\mathbf{h}^H(\omega)E\left[\mathbf{x}^{ab}(\omega)X_1^{a*}(\omega)\right] = E\left[X_1^{ab}(\omega)X_1^{a*}(\omega)\right] \quad (16)$$

$$|\mathbf{h}(\omega)|_1 \leq \rho\,|\mathbf{h}_{MVDR}(\omega)|_1 \quad (17)$$

Making ρ sufficiently small will cause some of the coefficients to be zero because of the nature of the ℓ_1-constraint (17). Thus, the ℓ_1 constraint does a kind of continuous subset selection. By increasing ρ to 1, we can incrementally select microphones with greater importance.

3.3. Incremental selection

Initialization Solve

$$\mathbf{h}_o(\omega_f) = arg\min \mathbf{h}^H(\omega_f)\mathbf{R}_V^{ab}(\omega_f)\mathbf{h}(\omega_f) \quad (18)$$

subject to

$$\mathbf{h}^H(\omega_f)E\left[\mathbf{x}(\omega_f)X_1^{a*}(\omega_f)\right] = E\left[X_1^a(\omega_f)X_1^{a*}(\omega_f)\right]. \quad (19)$$

Let set A and B be those of the microphones of directivities A and B that are deleted. We start from $Set\,A = \{\}$, $Set\,B = \{\}$ and $\rho = \rho_0$.

Step 1 Solve the problem

$$\mathbf{h}_\rho(\omega_f) = arg\min \mathbf{h}^H(\omega_f)\mathbf{R}_V^{ab}(\omega_f)\mathbf{h}(\omega_f) \quad (20)$$

subject to

$$\mathbf{h}^H(\omega_f)E\left[\mathbf{x}(\omega_f)X_1^{a*}(\omega_f)\right] = E\left[X_1^a(\omega_f)X_1^{a*}(\omega_f)\right] \quad (21)$$

$$|\mathbf{h}(\omega_f)|_1 \leq \rho\,|\mathbf{h}_0(\omega_f)|_1 \quad (22)$$

$$H_m^A(\omega_f) = 0 \ for \ m \in Set\,A \quad (23)$$

$$H_m^B(\omega_f) = 0 \ for \ m \in Set\,B. \quad (24)$$

The ℓ_1 constraint induces $h_\rho(\omega_f)$ sparse.

Step 2 If the problem of step 1 has no solution, increase ρ as $\rho \leftarrow \rho + \triangle\rho$, and go to step 1.

Step 3 When $|H_m^A(\omega_f)|_1 > |H_m^B(\omega_f)|_1$ and $H_m^A(\omega_f)$ is effective (e.g., $|H_m^A(\omega_f)|_1 > \delta|\mathbf{h}_0(\omega_f)|_1$), delete the mth microphone of type B by adding m to *Set B*. Similarly, when $|H_m^B(\omega_f)|_1 > |H_m^A(\omega_f)|_1$ and $H_m^B(\omega_f)$ is effective, delete the *m*th microphone of type A by adding *m* to *Set A*.

Step 4 If the total number of deleted microphones is $M-1$, exit the loop and adopt the directivity combination specified by *Set A* and *Set B*. Otherwise, increase ρ as $\rho \leftarrow \rho + \triangle\rho$, and go to step 1.

Note that the equality constraints of (23) and (24) can be removed by using the selection matrix $\mathbf{S}$, which omits the deleted microphone signals and reduces the dimension of the filter-coefficient vector $\mathbf{h}(\omega_f)$. The convex problem we should solve in step 1 can be rewritten as follows.

$$\tilde{\mathbf{h}}_\rho(\omega_f) = arg\min \tilde{\mathbf{h}}^H(\omega_f)\mathbf{S}\,\mathbf{R}_V(\omega_f)\,\mathbf{S}^H\,\tilde{\mathbf{h}}(\omega_f) \quad (25)$$

subject to

$$\tilde{\mathbf{h}}^H(\omega_f)\mathbf{S}E\left[\mathbf{X}(\omega_f)X_1^*(\omega_f)\right] = E\left[X_1(\omega_f)X_1^*(\omega_f)\right] \quad (26)$$

$$\left|\tilde{\mathbf{h}}(\omega_f)\right|_1 \leq \rho\,|\mathbf{h}_o(\omega_f)|_1 \quad (27)$$

3.4. Focusing

We use Doron's focusing technique, which can be applied to the array of an arbitrary geometry in the plane. Let the microphone's positions be given by (r_m, ϕ_m) in the polar coordinates, where its origin is set to the gravity center of these microphones.

The focusing matrix $\mathbf{T}(\omega_i)$ is given as

$$\mathbf{T}(\omega_i) = \mathbf{G}(\omega_f)\mathbf{G}^{\#}(\omega_i), \quad (28)$$

where $\mathbf{G}(\omega)$ is the following $2M \times (2N+1)$ matrix and $\mathbf{G}^{\#}(\omega) = \left[\mathbf{G}(\omega)\mathbf{G}^H(\omega)\right]^{-1}\mathbf{G}(\omega)$ is its generalized inverse matrix.

$$\mathbf{G}(\omega) = \begin{bmatrix} G_{1,-N}(\omega) & \cdots & G_{1,N}(\omega) \\ \vdots & & \vdots \\ G_{2M,-N}(\omega) & \cdots & G_{2M,N}(\omega) \end{bmatrix}. \quad (29)$$

Elements of $\mathbf{G}(\omega)$ are expressed as

$$G_{m,n}(\omega) = (j)^n J_n(k\,r_m)e^{j\,n\,\phi_m}, \quad (30)$$

where $J_n()$ is the nth-order Bessel function of the first kind and k is the wave number defined as $\omega/(sound\,velocity)$. We used $N = 12$ in the following simulations.

4. EVALUATION

We evaluated the performance of the proposed method by simulation. We used a linear microphone array of six pairs of closely spaced omnidirectional microphones, where the distance between the pairs was 2.5 cm and an inter-microphone distance inside a pair was 0.5 cm. From this pair, we generated hypercardioid directivity. An ominidirectional microphone in the first pair was used as the reference microphone. The proposed method was used to search for a better combination of omnidirectional and hypercardioid directivity for the remaining five microphone pairs. We used CVX, a package [13] [14] for solving the ℓ_1-constrained MVDR problem in the proposed method. Based on the estimated combination,

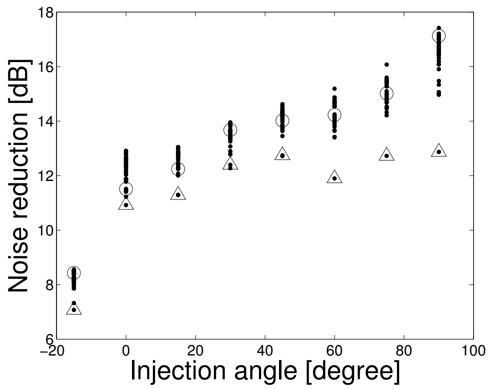

Fig. 2. Noise reduction performance for various injection angles: all microphone combinations ($\cdot$), identical omnidirecional microphones ($\triangle$) and combination found by proposed method ($\bigcirc$)

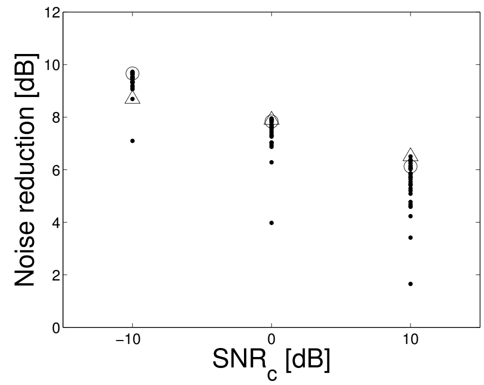

Fig. 3. Noise reduction performance for non-coherent plus coherent noise field ($SNR_c = -10$, 0, 10 dB): all microphone combinations ($\cdot$), identical omnidirecional microphones ($\triangle$) and combination by proposed method ($\bigcirc$)

conventional MVDR beamforming filters were computed at all frequencies.

We considered a planar configuration in which a target source, interference, and microphones are located on a single plane. The room size was $6 \times 7 \times 4$ m (length $\times$ width $\times$ height) and its reverberation time was 0.3 s. All room impulse responses were generated using "RIR generator" [15], which is based on the image-method proposed by Allen and Berkley [16] with some necessary modifications proposed by Peterson [17]. The sampling frequency was set to 16 kHz. The distance between the target source and the reference microphone was 2 m. The target source was speech-like noise (USASI) and its injection angle was -20 degrees. The noise consisted of pink noise. We used 512-point fast Fourier Transform for STFT, $\omega_f = 937.5$Hz, $\rho_0 = 0.3$, $\Delta\rho = 0.05$, and $\delta = 0.003$. Since large error occurs below the focusing frequency [7], we used 937.5–7125 Hz for the focusing of (6) and (7). We investigated the full-band noise-reduction factor, which is defined using signal to noise ratio (SNR) as

$$\frac{\text{full-band output SNR of MVDR beamformer}}{\text{full band input SNR of reference microphone}},$$

under the following two noise fields. In all situations, the proposed method was able to find the directivity combination within one or two executions of step 1.

Coherent noise field
This noise field was generated by an interference. The input SNR was set to 0 dB. Eight situations were simulated in which the injection angle of the interference was set to -15, 0, 15, 30, 45, 60, 75, or 90 degrees. Fig. 2 shows the full-band noise reduction factor of all $2^5 = 32$ microphone combinations ($\cdot$), that of omnidirecional microphones ($\triangle$), and that of the estimated combination ($\bigcirc$) for each inference injection angle. From this graph, the proposed method found a better combination than with an identical omnidirectional microphone array. Noise reduction was better for larger angular separation between the target source and an interference.

Non-coherent and coherent noise field
This noise field was generated by spatially-white (i.e., non-coherent) noises plus an interference. Input SNR of non-coherent noise was 20 dB and SNR of coherent noise (SNR_c) was set to -10, 0, and 10 dB. Fig. 3 shows the full-band noise reduction factor of all $2^5 = 32$ microphone combinations ($\cdot$), that of omnidirecional microphones ($\triangle$), and that of the estimated combination ($\bigcirc$) for each SNR_c. From this graph, the combination of an identical omnidirectional microphone array was almost best. The proposed method found a better combination at $SNR_c = -10$ dB and the combination very close to that of identical omnidirectional microphones at $SNR_c = 0$ and 10 dB

5. RELATION TO PRIOR WORK

The work presented here has focused on how to find a directivity combination by using variable selection induced by the ℓ_1 constraint. We expanded the idea of omnidirectional loudspeaker selection by least absolute shrinkage and selection operator (Lasso) [18] to MVDR beamforming, and used a different sparse constraint from [19]. This work is based on the idea of applying MVDR to focused PSD matrices [7] and the focusing technique [10][11].

6. CONCLUSION

We proposed a method for finding a microphone direcitivity combination for a nonidentical microphone array based on ℓ_1-based MVDR beamforming. Simulation results showed that the proposed method was able to find a better combination for a nonidentical microphone array in an efficient manner without examining all combinations.

7. REFERENCES

[1] H. L. V. Trees, *Detection, Estimation, and Modulation Theory*, Wiley, New York, 2002, IV, Optimum Array Processing.

[2] S. Gannot and I. Cohen, "Adaptive beamforming and postfiltering," in *Springer handbook of speech processing*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds. Springer-Verlag, New York, 2007.

[3] J. Benesty, J. Chen, and Y. Huang, *Microphone array signal processing*, Springer-Verlag, Berlin, 2008.

[4] E. Habets, J. Benesty, I. Cohen, S. Gannot, and J. Dmochowski, "New insights into the mvdr beamformer in room acoustics," *IEEE Trans. ASLP*, pp. 158–170, Jan. 2010.

[5] D. Levin, E. Habets, and S. Gannot, "Robust beamforming using sensors with nonidentical directivity patterns," in *Proc. ICASSP2013*, 2013, pp. 91–95.

[6] J. Cao, A. W. H. Khong, and S. Gannot, "On the performance of widely linear quaternion based mvdr beamformer for an acoustic sensor," in *Proc. IWAENC2014*, 2014.

[7] Y. Bucris, I. Cohen, and M. A. Doron, "Bayesian focusing for coherent wideband beamforming," *IEEE Trans. ASLP*, pp. 1282–1296, May 2012.

[8] E. D. Sena, H. Hacihabiboğlu, and Z. Cvetković, "On the design and implementation of higher order differential microphones," *IEEE Trans. ASLP*, pp. 162–174, Jan. 2012.

[9] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources," *IEEE Trans. ASSP*, pp. 823–831, Aug. 1985.

[10] M. A. Doron, E. Doron, and A. J. Weiss, "Coherent wide-band processing for arbitrary array geometry," *IEEE Trans. SP*, pp. 414–417, Jan. 1993.

[11] M. A. Doron and E. Doron, "Wavefield modeling and array processing, part ii-algorithms," *IEEE Trans. on SP*, pp. 2560–2570, Oct. 1994.

[12] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer, New York, 2001.

[13] Inc. CVX Research, "CVX: Matlab software for disciplined convex programming," http://cvx.com/cvx, Aug. 2012.

[14] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex program," in *Recent Advances in Learning and Control*, V. Blondel, S. Boyd, and H. Kimura, Eds., Lecture Notes in Control and Information Sciences, pp. 95–100. Springer-Verlag, 2008.

[15] E. Habets, "Room impulse response (rir) generator," http://home.tiscali.nl/ehabets/rirgenerator.html, July 2006.

[16] J. B. Allen and D. A. Berkley, "Image method for efficiently simu- lating small room acoustics," *J. Acoust. Soc. Amer.*, pp. 943–950, 1979.

[17] P. M. Peterson, "Simulating the response of multiple microphones to a single acoustic source in a reverberant room," *J. Acoust. Soc. Amer.*, pp. 1527–1529, Nov. 1986.

[18] G. D. Lilis, D. Angelosante, and G. B. Giannakis, "Sound field reproduction using lasso," *IEEE Trans. ASLP*, pp. 1902–1912, Nov. 2010.

[19] Y. Liu, Q. Wan, and X. Chu, "A robust beamformer based on weighted sparse constraint,," *Progress In Electromagnetics Research Letters*, pp. 53–60, 2010.