

# AD-HOC MICROPHONE ARRAY CALIBRATION FROM PARTIAL DISTANCE MEASUREMENTS

Mohammad J. Taghizadeh<sup>1,2</sup>, Afsaneh Asaei<sup>1</sup>, Philip N. Garner<sup>1</sup>, Hervé Bourlard<sup>1,2</sup>

<sup>1</sup>Idiap Research Institute, Martigny, Switzerland

<sup>2</sup>École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

{mohammad.taghizadeh, afsaneh.asaei, phil.garner, herve.bourlard}@idiap.ch

## ABSTRACT

We address the problem of ad hoc microphone array calibration where some of the distances between the microphones can not be measured. The conventional techniques require information about all the distances for accurate reconstruction of the array geometry. To alleviate this condition, we propose to exploit the properties of Euclidean distance matrices within the framework of low-rank matrix completion to recover the missing entries. We provide rigorous analysis to bound the calibration error using noisy measurements. This study elucidates the links between the performance and the structure of the missing distances, along with the size of the network. The experiments carried out on real data recordings demonstrate these theoretical insights. A significant improvement is achieved by the proposed Euclidean distance matrix completion algorithm over the state-of-the-art techniques for ad hoc microphone array calibration.

*Index Terms* – Ad-hoc microphone array calibration, Diffuse noise coherence model, Euclidean distance matrix completion

## 1. INTRODUCTION

Microphone arrays are widely used in distant audio technologies to enable source localization and separation [1–7], videoconferencing [8] and distant speech recognition in multiparty environments [9–12]. Ad hoc arrays provide a distributed and flexible infrastructure for high quality sound acquisition, and thus require to be calibrated to function in synergy. The focus of this paper is on the microphone array position calibration. This task is often achieved in two steps: estimation of the distances between the pairs of microphones and reconstruction of the array geometry from the pairwise distance information.

Prior art often considers activation of a (known) source signal in a specific configuration to estimate the distances between the microphones. Sachar et al. [13] used pulsed acoustic excitation of an array of five domed tweeters. The known test pulse is recorded to measure the transmit times between the speakers and microphones. A structured sparse coding approach is proposed in [14] to calibrate the microphones through a greedy procedure using an unknown signal at a given position and constructing an overcomplete dictionary of distant signals projected onto the array manifolds. Chen et al. [15] introduced an energy-based method for joint microphone calibration and speaker localization by computing the energy of the signal and

formulating a nonlinear optimization problem with a non-convex objective function to perform maximum likelihood estimation of the positions. Their approach requires at least five active sources. Recently, McCowan et al. [16] proposed to exploit the diffuse noise coherence model for computing the distances between the microphones. The coherence of the signals of the pairs of microphones is computed and fitted with a sinc function of their distance in a least square sense. This process is accomplished per signal frame; the frame-wise estimates are then combined using k-means clustering to obtain a single estimate.

Once the pairwise distances are estimated, the classic multi-dimensional scaling (MDS) method is used to reconstruct the microphone array geometry [17]. This method applies a double centering transformation to subtract the row and column means of the distance matrix. Then a low-rank projection is applied to extract the relative microphone positions. If some of the pairwise distances are missing, an extension of this method called MDS-MAP can be used. This method approximates the missing distances by the shortest path defined as the minimum sum of the distance measures of the constituent edges of the microphones. Recent advances incorporate the properties of the Euclidean distance matrix through definition of the appropriate cost functions via the algebraic s-stress method [18] and formulate effective optimization schemes for finding the geometry via semidefinite programming (SDP) [19].

In this paper, we use the coherence model of a diffuse field for estimating the distances between the microphones; the diffuse field has been shown to be a typical characteristic of the acoustic field in a reverberant enclosure [3]. The conventional approach provides reasonable estimates if the microphone array aperture size is less than 30 cm [16]. As the distance is increased, the frame-wise estimation leads to many outliers. Therefore, we apply an averaging of the frames' coherence prior to fitting a sinc function and propose a multi-stage clustering using k-means followed by GMM-based clustering to discard the outliers. This revised procedure enhances the reliable distance estimation up to 73 cm. The pairwise distances beyond that can not be estimated reliably, and thus assumed missing. To reconstruct the geometry, we exploit the recent advances in low-rank matrix completion and propose to recover the elements of the distance matrix via a Euclidean distance matrix completion algorithm. We provide rigorous analysis of the performance of the proposed approach to obtain the fundamental error bound and elucidate its relation to the problem setup. The proposed theory is demonstrated through the experiments conducted on real data recordings.

This paper is organized as follows. The procedure of pairwise distance estimation is explained in Section 2. The proposed Euclidean distance matrix completion is elaborated in Section 3. Theoretical insights of the calibration error is provided in Section 4. The

This work was supported by the Swiss National Science Foundation under the National Center of Competence in Research (NCCR) on “Interactive Multimodal Information Management” (IM2).

experimental evaluations are carried out in Section 5 and the conclusions are drawn in Section 6.

## 2. DISTANCE ESTIMATION

In this section, we summarize the theory of distance estimation based on the diffuse field coherence model and propose a multi-stage clustering approach to improve the accuracy and reliability.

### 2.1. Diffuse Field Coherence Model

A diffuse sound field consists of signals propagating with equal probability in all directions [20]. The coherence between the signals recorded with microphones  $i$  and  $k$  at angular frequency  $\omega$  is defined as

$$\Gamma_{ik}(\omega) = \frac{\Phi_{ik}(\omega)}{\sqrt{\Phi_{ii}(\omega)\Phi_{kk}(\omega)}}, \quad (1)$$

where  $\Phi_{ik}$  is the cross-spectral density between the signals of microphones  $i$  and  $k$ . Denoting the distance between the two microphones by  $d_{ik}$ , the coherence in a diffuse field is equivalent to  $\text{sinc}(\omega d_{ik}/c)$ , where  $c$  is the speed of sound. Based on this model, one can estimate the distance between each pair of microphones by fitting a sinc function to the coherence between their recorded signals [16]. In order to fit a sinc function to the coherence of the recorded signals, a simple quadratic objective function is used

$$\hat{d}_{ik}^j = \arg \min_d \left[ e_{ik}^j(d) = \sum_{\omega_{\min}}^{\omega_{\max}} \left| \Re\{\Gamma_{ik}^j(\omega)\} - \text{sinc}\left(\frac{\omega d}{c}\right) \right|^2 \right], \quad (2)$$

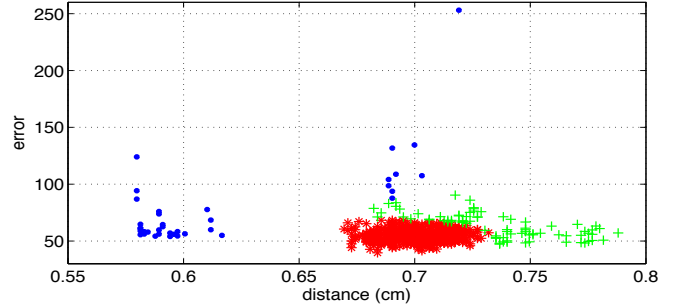
where  $[\omega_{\min}, \omega_{\max}]$  defines the frequency band in which the fitting is performed and  $j$  is the segment number in time; the  $\Re\{\cdot\}$  operator takes the real part of its input. The coherence of a segment is obtained by averaging the coherence over a sequence of frames [21]. The segment based estimates have to be combined to obtain a single estimation of the distance.

### 2.2. K-Means – GMM Clustering

Let  $J$  denote the number of segments used for extracting the distance between two microphones  $i$  and  $k$ . For each segment  $j$ ,  $(\hat{d}_{ik}^j, e_{ik}^j(\hat{d}_{ik}^j))$  represents the estimated distance and its corresponding objective value. To obtain a single estimate, k-means clustering has been employed and the mean of the cluster with a lower average objective value is used as an estimate of the distance [16].

In practice, the dependency of the level of diffuseness on frequency, and several physical parameters of the acoustic field, leads to many outliers in distance estimation, indicating a deviation from the model [22]; the number of these outliers increases as the distance grows rendering k-means clustering insufficient. Hence, we propose a multi-stage clustering by applying a GMM-based clustering to the result of k-means.

At the first stage, k-means is used to remove the cluster with an average larger error measure; an example is illustrated with blue points in Fig. 1. The winning cluster has many points conforming to similar distances and lower average error measure as illustrated with green crosses and red stars. To detect the outliers from the remaining points, we apply a GMM-based clustering and estimate the distance as the mean of the class with the least variance. Note that the outliers normally have a large spread in the two-dimensional feature space, that justifies the comparison of the classes based on their variance rather than the average objective value.



**Fig. 1:** Example of k-means followed by GMM clustering to remove the outliers for distance estimation. The blue dots have large errors, while the green crosses have more variance so the red stars are the winners.

In line with this heuristic justification, a 2-D histogramming approach is also effective for outlier removal and majority voting [21]; however, effective histogramming requires precise tuning of the bin resolutions whereas a two-stage k-means followed by GMM clustering yields more robust estimates. This approach enables us to estimate the distances less than  $d_{\max}$  and the ones beyond that are missing. We refer to them as structured missing entries.

## 3. GEOMETRY ESTIMATION

In this section, we propose a Euclidean distance matrix completion algorithm to recover the missing distances exploiting the low-rank characteristic of the squared distance matrix.

### 3.1. Euclidean Distance Matrix Completion

Matrix completion exploits the low-rank property to reconstruct a matrix given a subset of its entries. The distance matrix after applying a simple transformation has very low rank as stated through the following lemma [23].

**Lemma 1.** Consider a matrix  $M_{N \times N}$  consisting of the squared pairwise distances between  $N$  microphones defined as

$$M = [d_{ij}^2], \quad d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2, \quad i, j \in \{1, \dots, N\}$$

where  $d_{ij}$  is the distance between microphones  $i$  and  $j$  located at  $\mathbf{x}_i$  and  $\mathbf{x}_j$  in  $\mathbb{R}^c$ . The matrix  $M$  has rank at most  $\eta = c + 2$ .

Based on Lemma 1, there is a strong dependency among the entries of the squared distance matrix, thereby  $O(\eta N)$  measurements suffice to recover the missing components [24]. The matrix completion algorithm recovers  $M$  via the following optimization

$$\hat{M} = \arg \min_M \sum_{(i,j) \in E} (M_{ij} - \tilde{M}_{ij})^2 \quad (3)$$

subject to:  $\text{rank}(M) = \eta$ .

where  $E$  denotes the set of distances measured using the method explained in Section 2 and  $\tilde{M}$  is the corresponding measured squared distance matrix. In this paper, we use the procedure proposed by Keshavan et al. [24] for estimating a matrix given the desired rank.

The standard matrix completion algorithm recovers a low-rank matrix with elements as close as possible to the known entries. However, the recovered matrix does not necessarily correspond to a Euclidean distance matrix. Hence, we propose to project the matrix

obtained at each iteration of the matrix completion algorithm to the cone of Euclidean distance matrices,  $\text{EDM}^N$ . To this end, after one step of the gradient descent method, we apply a two-step projection,  $\mathcal{P} : \mathbb{R}^{N \times N} \mapsto \mathbb{S}_h^N \mapsto \text{EDM}^N$  to decrease the distance between the estimated matrix and the EDM cone. The  $\mathbb{S}_h^N$  designates the space of symmetric, positive hollow matrices. The projection onto  $\mathbb{S}_h^N$  is achieved by setting the diagonal and negative values to zero followed by averaging the symmetric elements; thereby, a new matrix  $\tilde{M}$  is obtained. The ultimate projection to  $\text{EDM}^N$  ensures that the matrix in addition to live in  $\mathbb{S}_h^N$ , has the property of  $-z^T \tilde{M} z \geq 0$  for all  $z \in \mathbb{R}^N$  such that  $\|z\| = 1$ ,  $\mathbb{1}_N^T z = 0$  where  $\mathbb{1}_N \in \mathbb{R}^N$  is the all ones vector and  $\cdot^T$  denotes the transpose operator [25]. To achieve the full EDM properties, we search in the EDM cone using the following cost function

$$\mathcal{H}(\mathbf{X}) = \left\| \mathbb{1}_N \mathbf{\Lambda}^T + \mathbf{\Lambda} \mathbb{1}_N^T - 2\mathbf{X}\mathbf{X}^T - \tilde{M} \right\|_F^2, \quad (4)$$

where  $\mathbf{X}_{N \times \kappa}$  denotes the position matrix whose  $i^{\text{th}}$  row,  $\mathbf{x}_i^T = [x_{i1}, \dots, x_{i\kappa}]^T$ , is the position of microphone  $i$  in  $\kappa$ -dimensional Euclidean coordinate and  $\mathbf{\Lambda} = (\mathbf{X} \circ \mathbf{X}) \mathbb{1}_\kappa$  where  $\circ$  denotes the Hadamard product. The minimum of  $\mathcal{H}(\mathbf{X})$  with respect to  $x_{i1}$  can be computed by equating the partial derivation of equation (4) to zero to obtain the new estimate  $\hat{\mathbf{X}}$  and the corresponding squared distance matrix  $\hat{M}$ . The stopping criterion is satisfied when the new estimate differs from the old one by less than a threshold or the maximum number of iterations is reached.

#### 4. THEORETICAL ANALYSIS

This section presents some theoretical insights on the performance of microphone array calibration using the matrix completion algorithm.

##### 4.1. Evaluation Measure

The error of geometry estimation has to be quantified with a cost function robust to rigid transformations (translation, rotation and reflection). We define the distance between  $\mathbf{X}$  and  $\hat{\mathbf{X}}$  as [26]

$$\text{dist}(\mathbf{X}, \hat{\mathbf{X}}) = \frac{1}{N} \left\| \mathbf{L} \mathbf{X} \mathbf{X}^T \mathbf{L} - \mathbf{L} \hat{\mathbf{X}} \hat{\mathbf{X}}^T \mathbf{L} \right\|_F, \quad (5)$$

$$\mathbf{L} = \mathbb{I}_N - (1/N) \mathbb{1}_N \mathbb{1}_N^T,$$

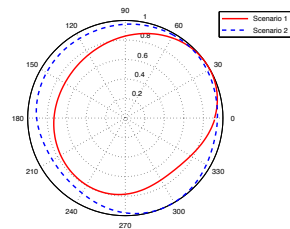
where  $\|\cdot\|_F$  denotes the Frobenius norm and  $\mathbb{I}_N$  is the  $N \times N$  identity matrix.

##### 4.2. Theoretical Guarantees

We denote the radius of a table where the microphones are deployed with  $a$ . The measurements of pairwise distances are noisy with an independent identically sub-Gaussian distribution of variance  $d_{ij}^2 \zeta^2$ . The constructed squared distance matrix has the smallest singular value designated as  $\sigma_\eta(\mathbf{M})$ .

The maximum distance that can be measured,  $d_{\max}$ , is assumed to be fixed and independent of  $N$ . In addition, we consider that some of the distances less than  $d_{\max}$  might not be accurately estimated, and thus missing with a probability  $1 - p$ . This consideration provides a better generalization for the results of our theoretical analysis.

Based on the following theorem, we guarantee that the calibration error has an upper bound that decreases as the local connectivity and size of the network grow.



**Fig. 2:** Diffuse noise recordings: Power pattern of the MVDR beamformer in (scenario 1) ambient noise and (scenario 2) augmented diffuse sound field.

**Theorem 1.** There exist constants  $C_1$  and  $C_2$ , such that the output  $\hat{\mathbf{X}}$  satisfies

$$\text{dist}(\mathbf{X}, \hat{\mathbf{X}}) \leq C_1 \frac{a^2}{p} + C_2 \zeta \frac{d_{\max}^2}{\sqrt{pN}} \quad (6)$$

with probability greater than  $1 - N^{-3}$ , provided that the right-hand side is less than  $\sigma_\eta(\mathbf{M})/N$ .

The proof of this theorem is explained in [27]. Based on Theorem 1, there is an upper bound on the error of microphone array calibration determined by two terms. The first term is indicated by the structured missing entries and the second term is indicated by the noisy estimates of the pairwise distances due to the deviation from a diffuse field model [27]. Theorem 1 elucidates that the error of calibration decreases as the connectivity of the network increases by adding further microphones. This effect has been investigated through the experimental evaluations presented in Section 5.

## 5. EXPERIMENTAL EVALUATION

This section presents the experimental analysis using real data recordings collected in the Idiapi smart meeting room.

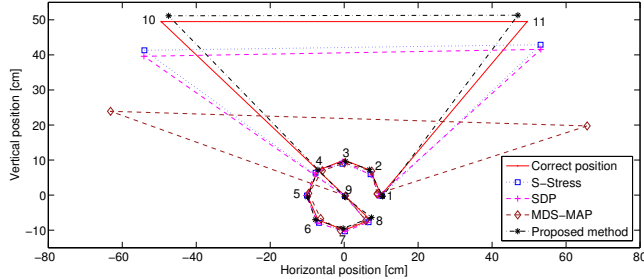
### 5.1. Recording Set-up

We consider a set-up of twelve microphones distributed on a planar area; eight of them are located on a circle with diameter 20 cm and one microphone is at the center. There are three additional microphones with 70 cm distance from the central microphone. This scenario mimics the MONC database [28]. The sampling rate is 48k while the processing applied for microphone calibration is based on a down-sampled signal of rate 16 k to reduce the computational cost of pairwise distance estimation.

The diffuse field signal is recorded in two scenarios: (1) collecting the ambient noise in the room without any additional source and (2) playing white Gaussian noise from two small loudspeakers put under the table and covering them with anti-acoustic material so that the direct path between the loudspeakers and the microphones are prohibited to ensure diffuseness [29]. A broadband minimum variance distortionless response (MVDR) beamformer [30] is used to scan the whole space and measure the power of the signal impinging from each direction. Fig. 2 shows the power patterns. We can see that a more isotropic (diffuse) field is achieved as the beamforming power pattern of the augmented sound field is closer to a circle.

### 5.2. Pairwise Distance Estimation

In order to estimate the pairwise distances, we take two microphone signals of length 30 s and window them into short frames of length 64 ms (i.e., 1024 samples) using a Tukey window (parameter



**Fig. 3:** Calibration of the eleven-element microphone array where several pairwise distances are missing. The geometries are estimated using MDS-MAP, SDP, S-stress and the proposed method.

= 0.25). For each five frames (segment), we compute the average coherence function and estimate the distance by fitting a sinc function as stated in (2). The segments are 80% overlapping, thereby 460 segments are obtained.

The next step to compute a single estimate of the distance between the two microphones is clustering the point estimates obtained per segment. The resulting segment-based estimates can generate many outliers partly due to the dependency of the diffuseness characteristics to the signal wavelength. The amount of error for various measured coherence functions along with the frequency dependencies are analyzed thoroughly in [22]. In a typical recording setup, more diffuseness is obtained at higher frequencies and the assumption of the model is violated at low frequencies. Boosting the power of the signal at high frequencies by augmenting the sound field as explained in Section 5.1, increases the power at high frequencies thus enhances the level of diffuseness. Detection of the outliers is further achieved by the multi-stage clustering approach elaborated in Section 2.2. Fig. 1 demonstrates the application of k-means followed by GMM clustering to improve the estimation of the distance between the microphones 9 and 11, cf. Fig. 3, with a distance of 70 cm. In the first stage, k-means clustering removes the segments that have larger errors (blue points). In the second stage, the green data points are removed using the GMM-based clustering. Using only k-means results in 71.2 cm distance while our approach yields 69.9 cm. This approach has been shown to enable reliable estimation of the pairwise distances up to 73 cm.

### 5.3. Microphone Calibration

We first evaluate estimation of the geometry when all of the pairwise distances are known, cf. Fig. 3-microphones one to nine. We use the recordings of the sound field in an almost silent room with (ambient noise) no activation of an additional noise. The calibration error,  $\text{dist}(\mathbf{X}, \hat{\mathbf{X}})$  is quantified based on (5). The calibration error for the circular array (microphones one to eight) using MDS, SDP, S-stress, matrix completion (MC) and the proposed Euclidean distance matrix completion (E-MC) algorithm are 10.49, 9.88, 12.12, 11.03 and 7.15 respectively. We repeat the same experiment using the recordings of an augmented diffuse field. The results are summarized in Table 1. We can see that boosting the power of the diffuse field enables more accurate calibration. The stopping criterion for the convergence of matrix completion is set to  $1e-4$ .

Although SDP yields the best performance in a weak sound field, by increasing the power of the diffuse sound, the s-stress method outperforms SDP. It can be justified as the s-stress is more likely to converge to a local optimum of its non-convex objective if the distance measures are noisier. By taking into account an additional microphone in the center, the local connectivity is increased

**Table 1:** Norm of error in squared distance ( $\text{cm}^2$ ) defined in (5) using different methods for ad hoc microphone array calibration.

	Known		Missing	
	8-mic	9-mic	11-mic	12-mic
MDS-MAP	9	8.13	434.4	472
SDP	9.09	8.63	141	135
S-Stress	6.86	6.14	125	95
MC	10.6	9.75	133	115
E-MC	6.5	5.85	49.6	46

which results in more accurate estimation of the distance matrix to perform calibration. These results demonstrate the theoretical insights provided in Section 4.2 as the performance improves with an additional microphone.

To evaluate the microphone calibration performance when some of the pairwise distances are missing, the positions of the nine closer microphones as estimated above are assumed fixed and used for calibration of the rest of the network. The proposed clustering technique enables reliable estimation of distances up to 73 cm. Hence, in our setup of 11 microphones,  $d_{10,11}$ ,  $d_{1,10}$ ,  $d_{8,10}$ ,  $d_{7,10}$ ,  $d_{5,11}$ ,  $d_{6,11}$  and  $d_{7,11}$  are missing. An extra microphone 12 is also included which is located with a symmetry to microphone 10. Hence,  $d_{12,11}$ ,  $d_{10,12}$ ,  $d_{3,12}$ ,  $d_{4,12}$ ,  $d_{5,12}$  are also missing. The calibration errors are listed in Table 1. The results show that considering further microphone improves the calibration performance which is in line with the theoretical analysis of Section 4.2.

We can see that the proposed Euclidean distance matrix completion algorithm achieves the best performance of microphone array geometry estimation. The worst result belongs to the MDS-MAP as the shortest path is a poor estimate of the missing entries. The s-stress and SDP search the Euclidean space corresponding to the feasible positions, hence their performance is more reasonable. The advantage of being constrained to a physically possible search space is incorporated in the Euclidean distance matrix completion algorithm and enables more robust calibration from partial noisy estimation of the distances. Once the distance matrix is recovered by classic matrix completion algorithms, MDS is used to find the coordinates of the microphones,  $\hat{\mathbf{X}}$ , whereas the proposed algorithm directly yields the coordinates. Fig. 3 illustrates an example of the estimated geometries using different approaches.

## 6. CONCLUSIONS

In this paper, we proposed a method for ad hoc microphone array calibration when only partial information about the pairwise distances can be measured. We exploited the diffuse field coherence model for estimation of the distances between adjacent microphones and proposed a multi-stage clustering scheme to detect and remove the outliers from the segment-wise estimates. The constructed distance matrix based on the coherence model is only partially revealed. To address the problem of missing distances, we proposed a Euclidean distance matrix completion algorithm and recovered a low-rank matrix where EDM properties are preserved. Moreover, we presented novel theoretical analysis of the calibration error where the maximum measurable distance is fixed irrespective of the size of the network and the measured distances have i.i.d. noises increasing with the distances. The experiments conducted on real data recordings demonstrate the applicability of the proposed method and the effectiveness of increasing the local connectivity and diffuse field power for accurate calibration.

## References

- [1] D. Malioutov, M. Çetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *Signal Processing, IEEE Transactions on*, vol. 53, no. 8, pp. 3010–3022, 2005.
- [2] M. S. Brandstein and H. F. Silverman, "A practical methodology for speech source localization with microphone arrays," *Computer Speech and Language*, vol. 11(2), 1997.
- [3] M. J. Taghizadeh, P. N. Garner, H. Bourlard, H. R. Abutalebi, and A. Asaei, "An integrated framework for multi-channel multi-source localization and voice activity detection," in *IEEE Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*, 2011.
- [4] J. H. DiBiase, "A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays," in *PhD Thesis, Brown University in Providence, Rhode Island, United States*, 1993.
- [5] A. Asaei, M. J. Taghizadeh, M. Bahrololoum, and M. Ghanbari, "Verified speaker localization utilizing voicing level in splitbands," *Signal Processing*, vol. 89(6), 2009.
- [6] H. T. Do, "Robust cross-correlation-based methods for sound-source localization and separation using a large-aperture microphone array," Ph.D. dissertation, Brown University, 2011.
- [7] A. Asaei, M. J. Taghizadeh, H. Bourlard, and V. Cevher, "Multi-party speech recovery exploiting structured sparsity models," in *12th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, 2011.
- [8] R. Cutler, Y. Rui, A. Gupta, J. Cadiz, I. Tashev, L.-w. He, A. Colburn, Z. Zhang, Z. Liu, and S. Silverberg, "Distributed meetings: a meeting capture and broadcasting system," in *Proceedings of the tenth ACM international conference on Multimedia*, 2002.
- [9] K. Kumatani, J. McDonough, and B. Raj, "Microphone array processing for distant speech recognition: From close-talking microphones to far-field sensors," *IEEE Signal Processing Magazine, Special Issue on Fundamental Technologies in Modern Speech Recognition*, 2012.
- [10] C.-T. Do, M. J. Taghizadeh, and P. N. Garner, "Combining cepstral normalization and cochlear implant-like speech processing for microphone array-based speech recognition," in *IEEE workshop on Spoken Language Technology (SLT)*, 2012.
- [11] A. Asaei, "Model-based sparse component analysis for multiparty distant speech recognition," Ph.D. dissertation, École Polytechnique Fédérale de Lausanne (EPFL), 2013.
- [12] M. Wolfel and J. McDonough, "Distant speech recognition," *New York: John Wiley & Sons*, 2009.
- [13] J. M. Sachar, H. F. Silverman, and W. R. Patterson, "Microphone position and gain calibration for a large-aperture microphone array," *IEEE Transactions on Speech and Audio Processing*, vol. 13(1), 2005.
- [14] A. Asaei, B. Raj, H. Bourlard, and V. Cevher, "Structured sparse coding for microphone array position calibration," in *Proceeding of 5th ISCA workshop on Statistical and Perceptual Audition, SAPA-SCALE Conference*, 2012.
- [15] M. Chen, Z. Liu, L. He, P. Chou, and Z. Zhang, "Energy-based position estimation of microphones and speakers for ad-hoc microphone arrays," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2007.
- [16] I. McCowan, M. Lincoln, and I. Himawan, "Microphone array shape calibration in diffuse noise fields," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16(3), 2008.
- [17] T. F. Cox and M. A. A. Cox, "Multidimensional scaling," *Chapman-Hall*, 2001.
- [18] A. Buja and D. F. Swayne, "Visualization methodology for multidimensional scaling," *Journal of Classification*, vol. 19, 2002.
- [19] P. Biswas, T. C. Liang, K. C. Toh, T. C. Wang, and Y. Ye, "Semidefinite programming approaches for sensor network localization with noisy distance measurements," *IEEE Transactions on Automation Science and Engineering*, vol. 3, 2006.
- [20] R. K. Cook, R. V. Waterhouse, R. D. Berendt, S. Edelman, and M. C. Thompson, "Measurement of correlations coefficients in reverberant sound fields," *Journal of the Acoustical Society of America*, vol. 27, 1955.
- [21] M. J. Taghizadeh, R. Parhizkar, P. N. Garner, and H. Bourlard, "Euclidean distance matrix completion for ad-hoc microphone array calibration," in *IEEE 18th International Conference in Digital Signal Processing*, 2013.
- [22] M. J. Taghizadeh, P. N. Garner, and H. Bourlard, "Enhanced diffuse field model for ad hoc microphone array calibration," *Signal Processing*, vol. 101, 2014.
- [23] P. Drineas, M. Javed, M. Magdon-Ismael, G. Pandurangant, R. Varrankoski, and A. Savvides, "Distance matrix reconstruction from incomplete distance information for sensor network localization," *Sensor and Ad Hoc Communications and Networks*, vol. 2, 2006.
- [24] R. H. Keshavan, A. Montanari, and S. Oh, "Matrix completion from a few entries," *IEEE Transactions on Information Theory*, vol. 56(6), 2010.
- [25] J. Dattorro, *Convex Optimization and Euclidean Distance Geometry*. USA: Meboo Publishing, 2012.
- [26] I. Borg and P. J. Groenen, "Modern multidimensional scaling theory and applications," in *Springer*, 2005.
- [27] M. J. Taghizadeh, "Theoretical analysis of euclidean distance matrix completion for ad hoc microphone array calibration," *Idiap, Idiap-RR Idiap-Internal-RR-06-2014*, 2014.
- [28] D. Moore and I. McCowan, "The multichannel overlapping numbers corpus," *Idiap resources available online*; <http://www.cslu.ogi.edu/corpora/monc.pdf>.
- [29] E. Habets, "Generating sensor signals in isotropic noise fields," *Journals of Acoustical Society of America*, vol. 122 (6), 2007.
- [30] M. J. Taghizadeh, P. N. Garner, and H. Bourlard, "Broadband beam pattern for multi-channel speech acquisition and distant speech recognition," in *IEEE, 7th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, 2012.