

# Euclidean Distance Matrix Completion for Ad-hoc Microphone Array Calibration

Mohammad J. Taghizadeh<sup>\*†</sup> Reza Parhizkar<sup>†</sup>, Philip N. Garner<sup>\*</sup> and Hervé Bourlard<sup>\*†</sup>

<sup>\*</sup>Idiap Research Institute, Martigny, Switzerland

<sup>†</sup>Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland

Email: {mohammad.taghizadeh, phil.garner, herve.bourlard}@idiap.ch, reza.parhizkar@epfl.ch

**Abstract**—This paper addresses the application of missing data recovery via matrix completion for audio sensor networks. We propose a method based on Euclidean distance matrix completion for ad-hoc microphone array location calibration. This method can calibrate a full network from partial connectivity information. The pairwise distances of microphones in close proximity are estimated using the coherence model of the diffuse noise field. The distance matrix of the ad-hoc network is constructed where the distances of the microphones above a threshold are missing. We exploit the low-rank property of the squared distance matrix and apply a matrix completion method to recover the missing entries. In order to constrain the Euclidean space geometry, we propose the additional use of the Cadzow algorithm for matrix completion. The applicability of the proposed method is evaluated on real data recordings where a significant improvement over the state-of-the-art is achieved.

**Index Terms**—Ad-hoc microphone calibration, Diffuse noise coherence, Matrix completion, Cadzow algorithm

## I. INTRODUCTION

Speech is one of the richest and most ubiquitous modalities of communication to enable human machine interaction. Microphone arrays provide a hands-free acquisition interface for the state-of-the-art technologies [1] and they have been applied in a broad range of applications including distant speech recognition [2, 3, 4], speaker localization [5] and speech separation [6, 7]. Recent advances in mobile computing and communication technologies suggest using cell phones, PDA's or tablets as a flexible acquisition set-up providing an ad-hoc network of microphones. However, the unknown prior information on relative locations of microphones is the first bottleneck to achieve high quality data processing. Microphone array calibration techniques aim to determine the geometry of the array. In the following, we overview previous approaches for calibration of the ad-hoc microphones to identify the practical challenges.

We can broadly group the prior art in three categories. The first group relies on transmitting a known signal to perform microphone calibration. Sachar et al. [8] presented an experimental setup using a pulsed acoustic excitation generated by five domed tweeters. The transmit times between speakers and microphones are used to find the relative geometry. In an alternative set-up, Raykar et al. [9] use a maximum length sequence or chirp signal in a distributed computing platform. The time difference of arrival of microphone signals are then computed by cross-correlation and used for estimating the

microphone locations. Since the original signal is known, these techniques are robust to noise and reverberation.

The second category enables using an unknown signal and the microphone calibration is usually integrated with source localization. Flanagan and Bell [10] propose a method using the Weiss-Friedlander technique where the sensor location and direction of arrival of the sources are estimated alternately until the algorithm is converged. Another approach is proposed in [11] by introducing an energy-based method for joint microphone-calibration and speaker-localization. The energy of the signal is computed and a nonlinear optimization problem is formulated to perform maximum likelihood estimation of the source-sensor positions. This method requires several active sources for accurate localization and calibration.

Recently, a third approach is proposed by McCowan and Lincoln [12] using the characteristics of a diffuse noise field model. A diffuse noise field is characterized by noise signals that propagate with equal probability in all locations and its coherence is defined by the sinc function of the distance of the two microphones. They propose to compute the distances by fitting the computed noise coherence with the sinc function in the least square sense. To increase the robustness, the noise frames are extracted and classified using k-means clustering.

The state-of-the-art techniques are usually applicable for conventional microphone arrays and estimation of the pairwise distances becomes noisy and unreliable as the distances between the microphones are increased. Thereby, some approaches are proposed to calibrate the full network of ad-hoc microphones given only partial information about the pairwise distances (e.g. MDS-MAP [13]). It has been shown that such methods can estimate the relative geometry of the ad-hoc microphone array even in the case that some of the pairwise distances are missing. However, most of these methods provide coarse approximations of the array geometry and their applicability is very limited.

This paper proposes to incorporate matrix completion constrained on Euclidean space properties for microphone array calibration. Matrix completion is a generic tool for estimating the unknown entries of a matrix from partial random information given prior knowledge about the (low) rank of the completed matrix. We first estimate the pairwise distances of the microphones in close proximity using the coherence characteristics of the signals of the two microphones in a diffuse noise field by an improved version of the method proposed in [12]; this approach implies a local connectivity constraint

as the pairwise distances of the further microphones can not be estimated. We show that a simple averaging among the noise frames yields more accurate estimates and speeds up the algorithm. We construct a matrix of all pairwise distances with missing entries corresponding to the unknown distances. We exploit the low-rank characteristics of the square of this matrix to enable estimation of the full-map geometry from a few known pairwise distances using matrix completion. To increase the accuracy, we incorporate properties of Euclidean Distance Matrices (EDM) in the matrix completion algorithm. We show that imposing EDM characteristics on matrix completion using the Cadzow algorithm improves the robustness and accuracy of extracting the ad-hoc microphone geometry.

The rest of the paper is organized as follows. In Section II, we explain how pairwise distances of the microphones are estimated using the coherence model of the diffuse noise field and how the proposed averaging method contributes to the accuracy of the method. In Section III, we describe the geometry estimation procedure from partial pairwise distances. The Euclidean matrix completion and the state-of-the-art s-stress method are compared and analyzed in Section IV. The conclusions are drawn in Section V.

## II. PAIRWISE DISTANCE ESTIMATION

We use the method proposed in [12] due to the practical assumptions of a diffuse noise model for audio applications [14] and no requirement for activating specific signals.

### A. Coherence of Distant Signals in a Diffuse Noise Field

We consider a scenario in which  $N$  microphones record a diffuse noise signal. Suppose that  $n_i$  and  $n_l$  represent the signal in time domain at microphones  $i$  and  $l$  respectively. The cross spectral density between  $n_i$  and  $n_l$  is

$$\Phi_{il}(\omega) = \hat{n}_i(\omega)\hat{n}_l^*(\omega),$$

where  $\omega$  denotes the frequency and  $*$  stands for conjugate transpose operator.  $\hat{n}_i(\omega)$  and  $\hat{n}_l(\omega)$  are Fourier transforms of noise signals  $n_i$  and  $n_l$  respectively. The coherence between noise signals is defined as

$$\Gamma_{il}(\omega) = \frac{\Phi_{il}(\omega)}{\sqrt{\Phi_{ii}(\omega)\Phi_{ll}(\omega)}} \quad (1)$$

For the diffuse noise field at each frequency component we have

$$\Gamma_{il}(\omega) = \text{sinc}\left(\frac{\omega d_{il}}{c}\right) \quad (2)$$

where  $d_{il}$  is the distance between the two microphones [15] and  $c$  is the speed of sound. Based on this model, estimation of distance between each two microphones is possible by fitting a sinc function on the coherence of their signal.

### B. Single-frame Distance Estimation

The following objective measure is proposed by McCowan and Lincoln [12] to fit a sinc function for a broadband

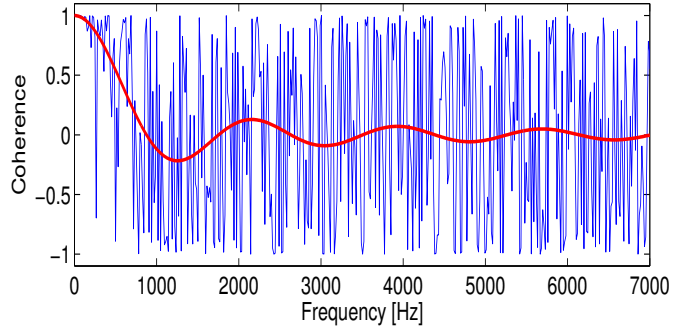


Fig. 1: Fitting a sinc function on one frame diffuse noise coherence; the correct distance is 20 cm and the estimated distance is 19.3 cm

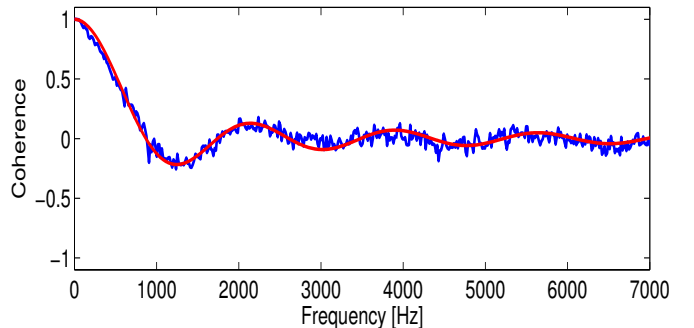


Fig. 2: Fitting a sinc function on average of 100 frames of diffuse noise coherence; The correct distance is 20 cm and the estimated distance is 19.8 cm

spectrum and estimate the pairwise distance. Fig. 1 represents an example of the coherence and fitted sinc function.

$$\delta_{il}^j(d) = \sum_{\omega=0}^{\omega_{max}} |\Re\{\Gamma_{il}^j(\omega) - \text{sinc}(\frac{\omega d}{c})\}|^2, \quad (3)$$

where  $\omega_{max}$  is chosen from the noise spectrum and is less than or equal to half of the sampling frequency, i.e.,  $2\pi f_s/2$  where  $j$  is the frame index. Minimizing  $\delta_{il}^j(d)$  over  $d$ , we obtain  $\tilde{d}_{il}^j$  estimation per frame as

$$\tilde{d}_{il}^j = \arg \min_d \delta_{il}^j(d) \quad (4)$$

Pairwise distance can be estimated from each frame of the noise signal.

### C. Averaging for Multi-frame Distance Estimation

Fig. 1, shows that it is difficult to find the relation between coherence of one frame and the sinc function. Hence, classification algorithms are proposed in [12] to remove the incorrect estimates and improve the performance by incorporating multiple frames. the classification stage is costly and requires long data segments to enable accurate classification.

We propose to average the coherence of multiple frames prior to fitting the sinc function. Fig. 2 demonstrates how averaging enables a more accurate least square regression. This method requires fewer number of frames than the classification

approach and it is very effective to improve the pairwise distance estimation performance.

To state it more precisely, we consider  $J$  frames to extract the distance between two microphones  $i$  and  $l$ . The averaging method is obtained via

$$\delta_{il}(d) = \sum_{\omega=0}^{\omega_{max}} \left| \left[ \frac{1}{J} \sum_{j=0}^{J-1} \Re \left( \Gamma_{il}^j(\omega) \right) \right] - \text{sinc} \left( \frac{\omega d}{c} \right) \right|^2, \quad (5)$$

$$\tilde{d}_{il} = \arg \min_d \delta_{il}(d)$$

### III. GEOMETRY ESTIMATION FROM PAIRWISE DISTANCES

In real scenarios, the estimated distances are noisy and in many situations computing them is impossible. For example, for the method explained in Section II-A, when the distance between two microphones is above a specific threshold ( $R$ ), the accuracy is very low. Thus, the distance matrix is noisy and is missing some of its entries. We propose to exploit the recently emerged matrix completion method [16, 17] for finding the missing entries of the distance matrix. In the following section, we focus on the matrix completion method and propose an additional variant for the particular application of microphone calibration.

#### A. Rank Property of Euclidean Distance Matrices

Consider a distance matrix  $\mathbf{D}_{N \times N}$  consisting of the pairwise distances between the  $N$  microphones constructed as

$$\mathbf{D} = [d_{ij}], \quad d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|, \quad i, j \in \{1, \dots, N\} \quad (6)$$

where  $d_{ij}$  is the distance between microphones  $i$  and  $j$  located at  $\mathbf{x}_i$  and  $\mathbf{x}_j$  on the plane. The matrix  $\mathbf{D}$  is symmetric and is often full rank.

The following lemma shows that a very simple transformation with saving information in the distance matrix  $\mathbf{D}$  can provide a very low rank matrix up to the space dimensionality.

**Lemma 1.** The squared distance matrix  $\mathbf{M}_{N \times N}$  constructed as  $\mathbf{M}_{ij} = [d_{ij}^2]$ , has rank at most  $\kappa + 2$  where  $\kappa$  denotes the dimension of the space where the microphones are deployed. For instance, if microphones are distributed on the plane or on a sphere,  $\mathbf{M}$  has rank 4 and if they are placed on a circle, the rank is 3. The proof is stated in [18].

This lemma shows that only partial distance measurements can characterize the whole matrix,  $\mathbf{M}$ . In other words, one can use only a few measurements to reconstruct the full matrix.

#### B. Euclidean Distance Matrix Completion

1) *Matrix Completion:* We recall our set-up that a few microphones are distributed in space with dimension  $\kappa$ . The distance matrix of microphone array is only partially available. The missing entries correspond to the microphone-pairs located greater than  $R$  meters apart. Our objective is to recover squared matrix  $\mathbf{M}_{N \times N}$  of rank  $\eta = \kappa + 2 < N$  from a sampling of its entries without having to ascertain all the  $N^2$  entries.

The approach proposed by *matrix completion* relies on the fact that a low-rank data matrix carries much less information

than number of its elements suggests. Intuitively, as the matrix  $\mathbf{M}$  has  $(2N - \eta)\eta$  degrees of freedom<sup>1</sup>, we need to know at least  $\eta N$  of the row entries as well as  $\eta N$  of the column entries reduced by  $\eta^2$  of the repeated values to recover a full  $\mathbf{M}$ .

Let  $E \subseteq [N] \times [N]$  denote the subset of known indices and  $\mathbf{M}^E$  the matrix consisting of the known entries of  $\mathbf{M}$  while filled with zeros as the unknown/missing entries.

$$\mathbf{M}_{i,j}^E = \begin{cases} \mathbf{M}_{i,j} & \text{if } (i, j) \in E, \\ 0 & \text{otherwise} \end{cases}, \quad (7)$$

which identifies a sampling set for  $\mathbf{M}$ . The matrix completion approach recovers full  $\mathbf{M}$  through the following optimization:

$$\begin{aligned} & \text{minimize} \quad \text{rank}(\widehat{\mathbf{M}}) \\ & \text{subject to} \quad \widehat{\mathbf{M}}_{i,j} = \mathbf{M}_{i,j} \quad (i, j) \in E. \end{aligned} \quad (8)$$

The matrix completion algorithm has three steps [17]: 1. Trimming, 2. Projection and 3. Minimizing the residual. In the trimming step, a row or a column is considered to be over-represented if it contains more samples than twice the average number of samples per row or column. These rows or columns can dominate the spectral characteristics of the observed matrix  $\mathbf{M}^E$ . Thus, some of their entries are removed uniformly at random from the observed matrix. Let  $\widetilde{\mathbf{M}}^E$  be the resulting matrix of this trimming step. In the second step, we first compute the singular value decomposition (SVD) of  $\widetilde{\mathbf{M}}^E$  as

$$\widetilde{\mathbf{M}}^E = \sum_{i=1}^N \sigma_i(\widetilde{\mathbf{M}}^E) \mathbf{u}_i \mathbf{v}_i^T \quad (9)$$

where  $\sigma_i(\cdot)$  denotes the  $i$ -th singular value of the matrix. Then, the rank- $\eta$  projection ( $\mathcal{P}_\eta(\cdot)$ ) returns the matrix obtained by setting to 0 all but the  $\eta$  largest singular values as

$$\mathcal{P}_\eta(\widetilde{\mathbf{M}}^E) = (N^2/|E|) \sum_{i=1}^{\eta} \sigma_i(\widetilde{\mathbf{M}}^E) \mathbf{u}_i \mathbf{v}_i^T = \mathbf{U}_0 \mathbf{S}_0 \mathbf{V}_0^T \quad (10)$$

Starting from the initial guess provided by the rank- $\eta$  projection  $\mathcal{P}_\eta(\widetilde{\mathbf{M}}^E)$ ,  $\mathbf{U} = \mathbf{U}_0$ ,  $\mathbf{V} = \mathbf{V}_0$  and  $\mathbf{S} = \mathbf{S}_0$ , the final step solves a minimization problem stated as follows. Given  $\mathbf{U} \in \mathbf{R}^{N \times \eta}$ ,  $\mathbf{V} \in \mathbf{R}^{N \times \eta}$  with  $\mathbf{U}^T \mathbf{U} = \mathbf{1}$  and  $\mathbf{V}^T \mathbf{V} = \mathbf{1}$ , define

$$F(\mathbf{U}, \mathbf{V}) = \min_{\mathbf{S} \in \mathbf{R}^{\eta \times \eta}} \mathcal{F}(\mathbf{U}, \mathbf{V}, \mathbf{S}), \quad (11)$$

$$\mathcal{F}(\mathbf{U}, \mathbf{V}, \mathbf{S}) = \frac{1}{2} \sum_{(i,j) \in E} (\mathbf{M}_{i,j} - (\mathbf{U} \mathbf{S} \mathbf{V}^T)_{i,j})^2 \quad (12)$$

$F(\mathbf{U}, \mathbf{V})$  is defined by minimizing the quadratic function  $\mathcal{F}$  over  $\mathbf{S}$ . The matrices  $\mathbf{U}$ ,  $\mathbf{V}$  will be estimated by minimizing the function  $F(\mathbf{U}, \mathbf{V})$  using gradient decent with line search. This last step tries to find the rank- $\eta$  matrix which has the closest values to  $\mathbf{M}$  on the measurement indices.

<sup>1</sup>The degrees of freedom can be estimated by counting the parameters in the singular value decomposition (the number of degrees of freedom associated with the description of the singular values and of the left and right singular vectors). When the rank is small, this is considerably smaller than  $N^2$  [19].

2) *Cadzow Projection to the EDM Properties*: The classic matrix completion algorithm as described above recovers a low-rank matrix which does not correspond to an Euclidean distance matrix. The Euclidean distance matrix is symmetric and the diagonal elements are zero. These properties are not incorporated in matrix completion algorithm. In order to obtain an Euclidean distance matrix completion, we propose a novel modification to the aforementioned algorithm in order to have matrices which are closer to an EDM.

To incorporate EDM properties, we apply a Cadzow-like method. The Cadzow (also known as Papoulis-Gershberg) algorithm [20] is a method for finding a signal which satisfies a composite of properties by iteratively projecting the signal into the property sets. We modify the matrix completion algorithm by inserting an extra step in the iterations of the algorithm. Recall that in the classic version of this algorithm a simple rank- $\eta$  approximation ( $\eta = 4$  for the distance matrix of a two-dimensional microphone array) is used as the starting point for the iterations using gradient descent on (12). After each step of the gradient descent, we apply the transformation  $\phi : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n}$  on the new matrix to make sure that the output satisfies the following properties sequentially:

- ◊ It is symmetric.
- ◊ It has a zero diagonal.
- ◊ It has only non-negative elements (we set all the negative elements to zero).
- ◊ It is rank- $\eta$ .

Thus, the modified iteration can be summarized in two steps:

- iteration  $k + 1/2$ :

$$\begin{aligned} \mathbf{U}^{k+1/2} &= \mathbf{U}^k + \mu \frac{\partial F(\mathbf{U}^k, \mathbf{V}^k)}{\partial \mathbf{U}} \\ \mathbf{V}^{k+1/2} &= \mathbf{V}^k + \mu \frac{\partial F(\mathbf{U}^k, \mathbf{V}^k)}{\partial \mathbf{V}} \\ \mathbf{S}^{k+1/2} &= \arg \min_{\mathbf{S}} \mathcal{F}(\mathbf{U}^k, \mathbf{V}^k, \mathbf{S}) \end{aligned}$$

- iteration  $k + 1$ :

$$(\mathbf{U}^{k+1}, \mathbf{V}^{k+1}, \mathbf{S}^{k+1}) = \phi(\mathbf{U}^{k+1/2}, \mathbf{V}^{k+1/2}, \mathbf{S}^{k+1/2}), \quad (13)$$

where  $\mu$  is the step size found using line search.

#### IV. EXPERIMENTAL ANALYSIS

In this section, we present experimental evaluations of the proposed theories. The microphone localization evaluation measure must be robust to rigid transformation. Hence, we use the distance between the actual locations  $\mathbf{X}$  and estimated locations  $\widehat{\mathbf{X}}$  as [21]

$$\begin{aligned} \text{dist}(\mathbf{X}, \widehat{\mathbf{X}}) &= \frac{1}{N} \|\mathbf{L}\mathbf{X}\mathbf{X}^T\mathbf{L} - \mathbf{L}\widehat{\mathbf{X}}\widehat{\mathbf{X}}^T\mathbf{L}\|_F, \\ \mathbf{L} &= \mathbf{I}_N - (1/N)\mathbf{1}_N\mathbf{1}_N^T, \end{aligned} \quad (14)$$

where  $\|\cdot\|_F$  denotes the Frobenius norm. The  $\mathbf{1}_N \in \mathbf{R}^N$  is the all ones vector and  $\mathbf{I}_N$  is the  $N \times N$  identity matrix and  $\mathbf{X}, \widehat{\mathbf{X}} \in \mathbf{R}^{N \times \kappa}$ . The distance measure stated in (14) is useful to compare the performance of different methods when the microphone array geometry is fixed.

#### A. Diffuse Noise Recording Set-up

We use the geometrical setup of the MONC corpus [22]. Twelve microphones are located in a planar area (i.e., two-dimensional space): 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. The microphones are Sennheiser MKE-2-5-C omnidirectional miniature lapel microphones. The floor is covered with carpet and surrounded with plaster walls having two big windows. The enclosure is a  $8 \times 5.5 \times 3.5 \text{ m}^3$  rectangular room and is moderately reverberant. It contains a centrally located  $4.8 \times 1.2 \text{ m}^2$  rectangular table. The sampling rate is 48 kHz.

In addition to the real recordings, we simulated the scenario described above to enable some evaluations in a controlled set-up. The results on the simulated data indicate the performance bound of the methods. We consider 32 white Gaussian noise sources distributed in the room. The room impulse responses are generated with the image source model [23] using intra-sample interpolation up to 15<sup>th</sup> order reflections. The corresponding reflection ratio,  $\beta$  used by the image model was calculated via Eyring's formula:

$$\beta = \exp(-13.82/[c \times (L_x^{-1} + L_y^{-1} + L_z^{-1}) \times T]), \quad (15)$$

where  $L_x, L_y$  and  $L_z$  are the room dimensions,  $c$  is the speed of sound in the air and  $T$  is the room reverberation time. In our experiments,  $T = 300 \text{ ms}$  and the direct-path propagation is discarded from the impulse response for generating a diffuse noise field [24].

#### B. Geometry Estimation on Simulated Data

In order to estimate the pairwise distances, we take two microphone signals of length 2.14 s and frame them into short windows of length 1024 samples using a Tukey function (parameter = 0.25) and apply Fourier transform. For each frame, we compute the coherence function through (1). The average of the coherence functions are computed and used for estimation of the pairwise distance by fitting a sinc function as stated in (5). This procedure is repeated for all microphone pairs to construct an estimated distance matrix. The geometry of the array is extracted using the state-of-the-art s-stress method by solving the following optimization problem [21]

$$\widehat{\mathbf{X}} = \arg \min_{\mathbf{X}} \sum_{(i,j) \in E} \left( \|\mathbf{x}_i - \mathbf{x}_j\|^2 - \widehat{d}_{ij}^2 \right)^2 \quad (16)$$

This method is a robust and accurate localization technique where the search space is constrained to the Euclidean geometry. Fig. 3 illustrates the microphone calibration results; The reconstruction error is 1.43 as defined in (14). The microphone calibration error using the baseline method proposed in [12] is estimated as 4. In addition, our averaging method speeds up the calibration by a factor 60 compared to [12] since the k-means clustering is not required to identify the accurate frames.

To perform evaluations for a scenario having microphones at long distances, we consider the three additional microphones. The proposed method enables a reasonable esti-

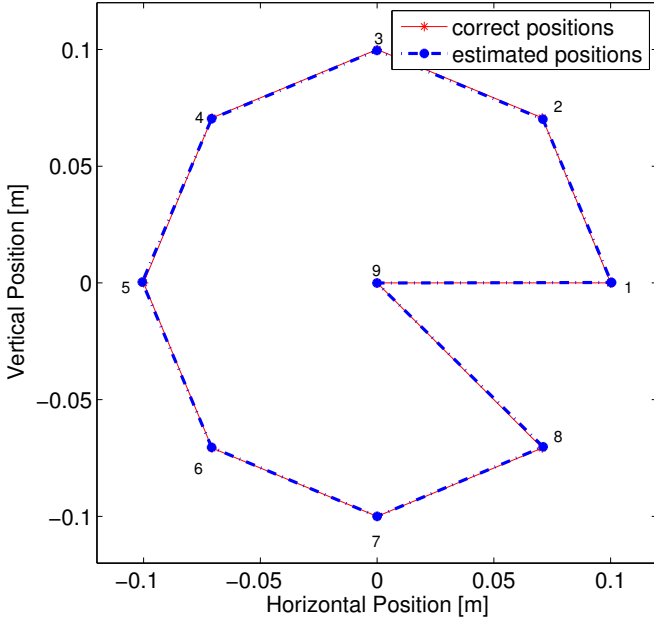


Fig. 3: Calibration of a 9-element microphone array with synthetic diffuse noise. The distances are estimated using averaging method and localization is obtained by the s-stress algorithm.

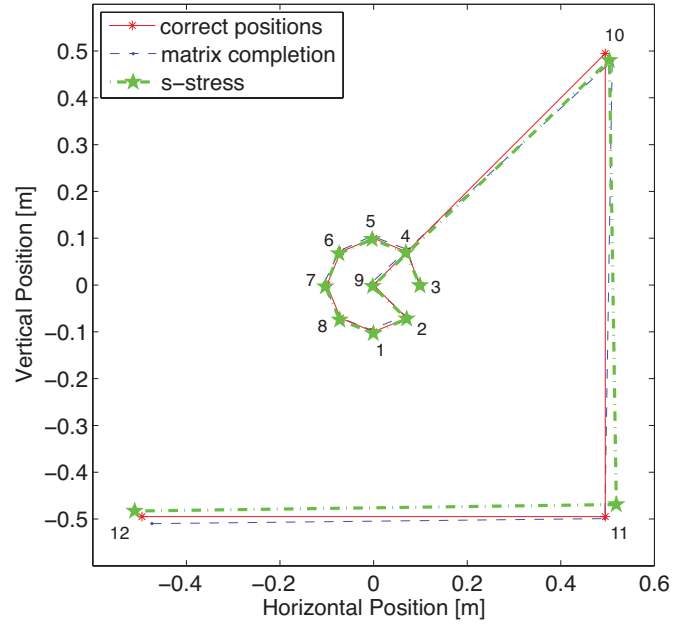


Fig. 4: Calibration of a 12-element microphone array on synthetic diffuse noise. First the central microphones are localized using the s-stress and then either s-stress or matrix completion is used to calibrate the full network.

mate up to 73 cm; we confirm empirically that the distances beyond that are not reliably estimated so we regard them as *missing*. Thereby, the following entries of the Euclidean distance matrix are missing,  $d_{10,11}$ ,  $d_{10,12}$ ,  $d_{1,10}$ ,  $d_{7,10}$ ,  $d_{8,10}$ ,  $d_{12,11}$ ,  $d_{5,11}$ ,  $d_{6,11}$ ,  $d_{7,11}$ ,  $d_{3,12}$ ,  $d_{4,12}$ ,  $d_{5,12}$  (see Fig. 4). Microphone calibration is achieved in two steps. In the first step, s-stress method is used to find 9 close microphones. These microphones are then used as anchor points. In the second step, either s-stress or matrix completion method is used to calibrate the full network. Fig. 4 demonstrates the results. The estimated error for matrix completion and s-stress methods are 33 and 65 respectively.

### C. Geometry Estimation on Real Data

In the first step, the geometry of the 9 microphones are estimated. To further improve the performance, we use a two-dimensional (errors vs. pairwise distances) histogram to remove the outliers; the resolution of the bins is chosen to be 2.5 mm. The two-dimensional histogram is very fast and it outperforms k-means clustering. The *green* curve in Fig. 5 depicts the results based on our averaging method on 100 frames; the estimated error is 8.04. The *blue* curve shows the improved estimates by the hybrid of averaging method and outlier detection using a 1000 frames where the averaging method is applied on five frames to estimate the pairwise distances and construct the two-dimensional histogram; the estimated error is 5.

To estimate the full network geometry where the pairwise distances of far microphones are missing, we apply the proposed Euclidean distance matrix completion technique and compare the results with the alternative s-stress

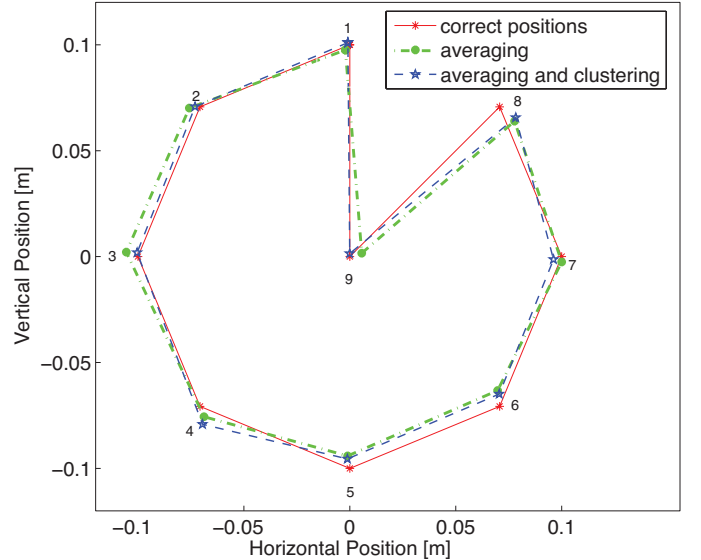


Fig. 5: Calibration of a 9-element microphone array on real diffuse noise recordings using averaging and a hybrid of averaging and histogram-based clustering. Localization is obtained by the s-stress method.

method. Fig. 6 demonstrates the results; similar to the previous test,  $d_{10,11}$ ,  $d_{10,12}$ ,  $d_{1,10}$ ,  $d_{7,10}$ ,  $d_{8,10}$ ,  $d_{12,11}$ ,  $d_{5,11}$ ,  $d_{6,11}$ ,  $d_{7,11}$ ,  $d_{3,12}$ ,  $d_{4,12}$ ,  $d_{5,12}$  are missing from the distance matrix. We observe that the matrix completion method combined with Cadzow algorithm yields the best results with error 54. The s-stress and the classic matrix completion methods have error of 99 and 159 respectively. Table I summarizes results for a



simulated scenario and a real scenario in a large configuration with 12 microphones. In the simulated scenario in which we can control diffuseness in the room, the error in estimation of distances is lower than the real scenario and the final geometry estimations are more accurate. Furthermore, integration of the matrix completion and Cadzow method can highly improve the results of geometry estimation especially in real scenario.

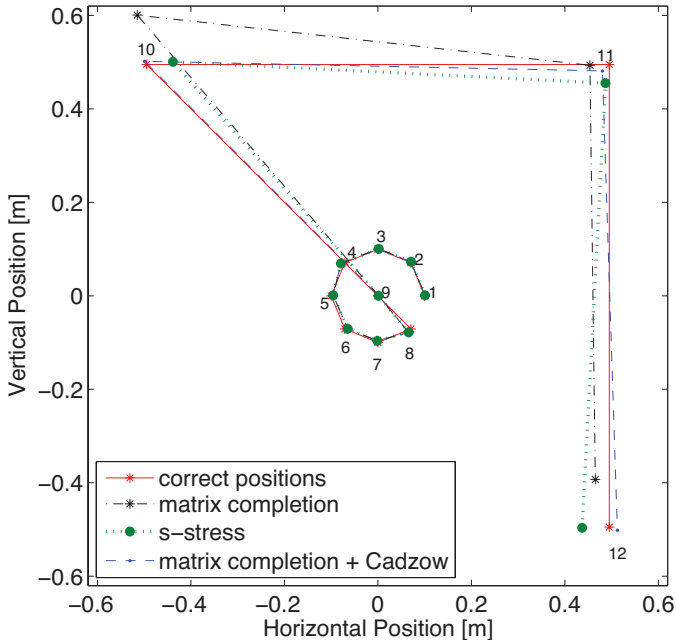


Fig. 6: Calibration of a 12-element microphones array in real diffuse noise. The central array is calibrated using the s-stress method. Then different approaches are used to calibrate the full network while several pairwise distances are missing.

Methods/Scenarios	Simulation	Real data
Matrix completion	33	159
S-Stress	65	99
Matrix completion+cadzow	25	54

TABLE I: Performance comparison of different approaches for extraction of the microphone array geometry

## V. CONCLUSIONS

We studied calibration of an arbitrary-sized microphone array using Euclidean distance matrix completion technique. The partial information about the pairwise distances is obtained exploiting the coherence model of the diffuse noise field. The calibration of the full network from partial known entries of the distance matrix is achieved using the Euclidean distance matrix completion method where Cadzow algorithm is used to impose the EDM properties. The evaluations conducted on real data recordings demonstrate the effectiveness of the proposed method and it outperforms the standard matrix completion and s-stress techniques. Although in low noise scenario (i.e. synthetic data) the standard matrix completion performs better than s-stress, it is highly sensitive to the noisy entries that we obtain in real scenarios and the projection on the

Euclidean distance matrix property sets enables a significant improvement. As a future plan we are interested to extend the simulated and real scenarios to complete ad-hoc geometry, and also extract results for other matrix completion methods.

## ACKNOWLEDGMENT

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

## REFERENCES

- [1] M. J. Taghizadeh, P. N. Garner, and H. Bourslard, “Broadband beam pattern for multi-channel speech acquisition and distant speech recognition,” in *Submitted to the 7th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, 2012.
- [2] A. Asaei, H. Bourslard, and V. Cevher, “Model-based compressive sensing for distant multi-party speech recognition,” in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2011.
- [3] 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.
- [4] A. Asaei, P. N. Garner, and H. Bourslard, “Sparse component analysis for speech recognition in multi-speaker environment,” in *Proceedings of INTERSPEECH*, 2010.
- [5] A. Asaei, M. J. Taghizadeh, M. Bahrololum, and M. Ghanbari, “Verified speaker localization utilizing voicing level in split-bands,” *Signal Processing*, vol. 89(6), 2009.
- [6] A. Asaei, M. J. Taghizadeh, H. Bourslard, and V. Cevher, “Multi-party speech recovery exploiting structured sparsity models,” in *12th Annual Conference of the International Speech Communication Association (INTER-SPEECH)*, 2011.
- [7] A. Asaei, M. Golbabaee, H. Bourslard, and V. Cevher, “Structured sparsity models for multiparty speech recovery from reverberant recordings,” <http://arxiv.org/abs/1210.67666>, 2012.
- [8] 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.
- [9] V. C. Raykar, I. V. Kozintsev, and R. Lienhart, “Position calibration of microphones and loudspeakers in distributed computing platforms,” *IEEE Transactions on Speech and Audio Processing*, vol. 13(1), 2005.
- [10] B. P. Flanagan and K. L. Bell, “Array self-calibration with large sensor position errors,” *Signal Processing*, vol. 81, 2001.
- [11] 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.

- [12] 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.
- [13] Y. Shang, W. Ruml, Y. Zhang, and M. P. J. Fromherz, "Localization from connectivity in sensor networks," *IEEE Transactions on Parallel Distributed Systems*, vol. 15(11), 2004.
- [14] M. J. Taghizadeh, P. N. Garner, H. Bourslard, 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.
- [15] B. Rafaely, "Spatial-temporal correlation of a diffuse sound field," *Journal of the Acoustical Society of America*, vol. 107, 2000.
- [16] E. J. Candes and Y. Plan, "Matrix completion with noise," *IEEE Signal Processing Magazine*, vol. 98(6), 2010.
- [17] R. H. Keshavan, A. Montanari, and S. Oh, "Matrix completion from a few entries," *IEEE Transactions on Information Theory*, vol. 56(6), 2010.
- [18] P. Drineas, M. Javed, M. Magdon-Ismael, G. Panduranga, R. Verrankoski, and A. Savvides, "Distance matrix reconstruction from incomplete distance information for sensor network localization," *Sensor and Ad Hoc Communications and Networks*, vol. 2, 2006.
- [19] E. Candes and B. Recht, "Exact matrix completion via convex optimization," *Magazine Communications of the ACM*, vol. 55, 2012.
- [20] J. Cadzow, "Signal enhancement-a composite property mapping algorithm," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, 1988.
- [21] I. Borg and P. J. Groenen, "Modern multidimensional scaling theory and applications," in *Springer*, 2005.
- [22] D. Moore and I. McCowan, "The multichannel overlapping numbers corpus," *Idiap resources available online*, <http://www.cslu.ogi.edu/corpora/monc.pdf>.
- [23] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of Acoustic Society of America*, vol. 65, 1979.
- [24] E. Habets, "Generating sensor signals in isotropic noise fields," *Journals of Acoustical Society of America*, vol. 122 (6), 2007.