RELATIVE TRANSFER FUNCTION MODELING FOR SUPERVISED SOURCE LOCALIZATION

Bracha Laufer[1], *Ronen Talmon*[2] *and Sharon Gannot*[1]

[1] Faculty of Engineering
Bar-Ilan University
Ramat-Gan, 52900, Israel
bracha_gold@walla.com, Sharon.Gannot@biu.ac.il

[2] Department of Mathematics
Yale University
New Haven, CT 06520-8283, USA
ronen.talmon@yale.edu

ABSTRACT

Speaker localization is one of the most prevalent problems in speech processing. Despite significant efforts in the last decades, high reverberation level still limits the performance of localization algorithms. Furthermore, using conventional localization methods, the information that can be extracted from dual microphone measurements is restricted to the time difference of arrival (TDOA). Under far-field regime, this is equivalent to either azimuth or elevation angles estimation. Full description of speaker's coordinates necessitates several microphones. In this contribution we tackle these two limitations by taking a *manifold learning* perspective for system identification. We present a training-based algorithm, motivated by the concept of *diffusion maps*, that aims at recovering the fundamental controlling parameters driving the measurements. This approach turns out to be more robust to reverberation, and capable of recovering the speech source location using merely two microphones signals.

Index Terms— acoustic source localization, manifold learning, diffusion kernel, relative transfer function.

1. INTRODUCTION AND MOTIVATION

The problem of source localization has drawn the attention of many researchers during the last decades. Many contributions adopt a dual-stage approach. In the first stage, the TDOA is estimated based on microphone pairs [1]. In the second stage, TDOA readings from several microphone pairs are combined to localize the speaker. Traditional methods are incapable of fully localizing the coordinates of a speaker from a single microphone pair measurements. Of special interest are TDOA estimation methods which are based on blind channel identification techniques [2, 3]. In this family of methods, a main peak of the identified acoustic path is taken as the TDOA estimate. One fundamental disadvantage of the traditional methods is a severe performance degradation in reverberant environments. Due to the reverberation phenomenon, most TDOA estimators exhibit multiple-peak output corresponding to the direct arrival and secondary arrivals. In low direct to reverberant ratio (DRR) conditions, it is usually not guaranteed that the highest peak corresponds to the direct arrival.

Talmon et al. [4] introduced a method based on *manifold learning* paradigm, and in particular *diffusion kernels*. The main essence of this approach, compared with traditional system identification methods, is that it aims at specifying the fundamental controlling parameters of the acoustic impulse response (AIR). Rather then fitting the identified system to a predefined model, the new approach inspects the underlying, latent, parameters of the system through a training set, given in advance. Since, assumably, the position of the source is the only varying degree-of-freedom of the system at hand, this process is capable of recovering the unknown source location. The key point of the algorithm is to use an appropriate diffusion kernel with a specifically-tailored distance measure, that is capable of finding the underlying independent parameters, dominating the system. Talmon et al. [5] have applied this method to a single microphone system with a white Gaussian noise (WGN) input. In this setting, the differences between the AIRs are attributed only to the source location, since the input signal is assumed to be stationary.

In the current contribution we adopt this paradigm and adapt it to a more realistic setting where the source is a speech signal rather than a WGN signal in the original contribution. The power spectral density of the speech signal is non-flat (as well as non-stationary). Hence, the spectral variations may blur the variations attributed to the different possible locations of the source.

In order to mitigate this problem, we conduct two major changes in the algorithm presented in [5]: 1) a second microphone is added and 2) the feature vector, that was originally based on the correlation function is replaced by power spectral density (PSD)-based vector. Accordingly, the relative transfer function (RTF) relating the microphone pair, consists of only the relevant information for generating a diffusion kernel (and the irrelevant source PSD is cancelled out). This is the key point in revealing the source location accurately when the input is a speech signal.

In addition, whereas in the traditional method only the TDOA can be revealed using two microphones, the proposed method is able to discover the exact location. Our algorithm recovers the underling parameters which control the system, therefore is not limited to one dimension only, but rather to the real dimensions of the system. In the experimental part we demonstrate this property by extracting both the azimuth and elevation angles, while the radius remains constant. 2-D sound localization, in the binaural hearing context, was presented in [6].

2. PROBLEM FORMULATION

A source is located in a reverberant enclosure and its speech is picked up by a microphone pair. The goal of the proposed algorithm is to localize the speaker from the measured microphone signals.

Consider a set of M (unknown) source locations, $\boldsymbol{\theta}_i = [\phi_i, \theta_i, \rho_i];\ i = 1, \ldots, M$. ϕ_i and θ_i are the azimuth and elevation angles and ρ_i is the distance of a source at the ith location, measured with respect to one of the microphones. Let $h_{\boldsymbol{\theta},ki}(n)$, $k = \{1, 2\}$ denote the AIR relating the source at location $\boldsymbol{\theta}_i$ and

978-1-4799-0972-8/13/$31.00 ©2013IEEE

microphone k. It will be shown in Sec. 3 that the algorithm is capable of localizing several *nonconcurrent* speakers. For the single source case we can assume $M = 1$.

The AIR fully characterizes the acoustic scenario. We will show that for a *fixed* acoustic environment (i.e. room characteristics and microphone constellation are unchanged during the experiments), the only degrees-of-freedom controlling the AIRs are the source coordinates.

The received microphone signals are given by:

$$y_{ki}(n) = h_{\boldsymbol{\theta},ki}(n) * x_i(n) \tag{1}$$

where $x_i(n)$ and $y_{ki}(n)$ are the input and output signals, respectively, corresponding to a *test set* of M unknown parameters $\Theta = \{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_M\} \subset \mathbb{R}^d$, specifying M locations of the sources. Let d be the dimension of the parameter vector, where, $d = 3$ in the aforementioned 3-D case, but as detailed in the experimental study in Sec. 4, it can be also set to $d = 1$ or $d = 2$, depending on the scenario.

According to the supervised learning paradigm, we wish to recover these test parameters using only the microphone (measured) signals and a training set that is available beforehand. As a training data we use *labeled* microphone signals produced by a source at a set of m predefined locations $\bar{\Theta} = \{\bar{\boldsymbol{\theta}}_1, \dots, \bar{\boldsymbol{\theta}}_m\} \subset \mathbb{R}^d$. In order to generate the training set, we play from each location in $\bar{\Theta}$ an arbitrary WGN signal of a finite length, and record the received signals in the microphone pair:

$$\bar{y}_{ki}(n) = h_{\bar{\boldsymbol{\theta}},ki}(n) * x_i(n). \tag{2}$$

For each location in the training set, we collect L additional measurements, originating from slightly perturbed locations in the vicinity of $\bar{\Theta}$. Let $\{\boldsymbol{\theta}_{i_j}\}_{j=1}^{L}$ denote the small perturbations of $\bar{\boldsymbol{\theta}}_i$, and $\{x_{i_j}(n), y_{ki_j}(n)\}_{j=1}^{L}$ denote the corresponding input and output signals, associated with this parameter. It is important to emphasize that the input signal for the training set is a WGN, while our goal is to localize a speech source. The flat frequency content of a WGN signal makes it a better training signal, due to its ability to fully excite the frequency response of the AIR.

3. THE PROPOSED ALGORITHM

We propose a *parameter inference* method that extracts only the information relevant for location estimation and is indifferent to other nuisance factors.

3.1. The Feature Vector

In the first stage of the proposed algorithm a feature vector is extracted from the received data. Talmon et al. [5] selected a feature vector comprised of several lags of the auto-correlation function of the received signals, $c_{y_i}(\tau) = h_{\boldsymbol{\theta}_i}(\tau) * h_{\boldsymbol{\theta}_i}(-\tau) * c_{x_i}(\tau)$. While this is a reasonable choice for a WGN input, since its correlation function is $c_{x_i}(\tau) = \delta(\tau)$, using this feature with an arbitrary speech input, will mask the desired controlling parameters, namely the source location, due to the spectral changes of the input signal. Hence, we propose to replace the auto-correlation by the cross-PSD (CPSD), utilizing the recordings from both microphones. Using the CPSD, the RTF relating the two microphones can be straightforwardly estimated:

$$\begin{aligned} T_{y_{1i}y_{2i}}(e^{i\omega_r}) &= \frac{S_{y_{1i}y_{2i}}(e^{i\omega_r})}{S_{y_{2i}y_{2i}}(e^{i\omega_r})} \\ &= \frac{S_{x_ix_i}(e^{i\omega_r}) H_{\boldsymbol{\theta}_{1i}}(e^{i\omega_r}) H^*_{\boldsymbol{\theta}_{2i}}(e^{i\omega_r})}{S_{x_ix_i}(e^{i\omega_r}) |H_{\boldsymbol{\theta}_{2i}}(e^{i\omega_r})|^2} = \frac{H_{\boldsymbol{\theta}_{1i}}(e^{i\omega_r})}{H_{\boldsymbol{\theta}_{2i}}(e^{i\omega_r})} \end{aligned} \tag{3}$$

where $S_{y_{1i}y_{2i}}(e^{i\omega_r})$ is the CPSD between $y_{1i}(n)$ and $y_{2i}(n)$, $S_{y_{2i}y_{2i}}(e^{i\omega_r})$ is the PSD of $y_{2i}(n)$ and $S_{x_ix_i}(e^{i\omega_r})$ is the PSD of the source located at $\boldsymbol{\theta}_i$, $H_{\boldsymbol{\theta}_{1i}}(e^{i\omega_r})$ and $H_{\boldsymbol{\theta}_{2i}}(e^{i\omega_r})$ are the acoustic transfer functions (ATFs) of the respective AIRs, and $\omega_r = \frac{2\pi r}{D}$; $r = 0, \dots, D-1$ denotes a discrete frequency index.

Note, that since the speech PSD is cancelled out, the RTFs solely depend on the characteristics of the AIRs. Let $\mathbf{T}_i \subset \mathbb{R}^D$ denote the RTF feature vector of the test set $\{y_{1i}(n), y_{2i}(n)\}$. It is comprised of D frequency bins of the RTF $\{T_{y_{1i}y_{2i}}(e^{i\omega_r})\}_{r=0}^{D-1}$. Respectively, $\overline{\mathbf{T}}_i \subset \mathbb{R}^D$ denotes the RTF feature vector of the training set $\{\bar{y}_{1i}(n), \bar{y}_{2i}(n)\}$.

The vectors $\{\mathbf{T}_{i_j}\}_{j=1}^{L}$ of the perturbed measurements, can be viewed as a *cloud of points* around $\overline{\mathbf{T}}_i$ in $\mathbb{R}^D$, and can be therefore utilized to estimate the local covariance matrix of $\overline{\mathbf{T}}_i$:

$$\hat{\bar{\boldsymbol{\Sigma}}}_i = \frac{1}{L} \sum_{j=1}^{L} \mathbf{T}_{i_j} \mathbf{T}_{i_j}^T. \tag{4}$$

This covariance matrix is a key point in recovering the inherent structure of the data, as it uses relationships in the parametric space to infer meaningful distances in the observable space. It will be the basis for defining an appropriate Mahalanobis distance used for constructing the diffusion kernel in the sequel. This covariance matrix estimates a local Jacobian-based metric distortion of the transformation that maps the parameter space (location parameters in our case) into the observable space (RTF values estimated from measured microphone signals). The reader is referred to [7] for more details.

3.2. Training Stage

An *affinity matrix* $\mathbf{W}$ between the m training samples in $\bar{\Theta}$ can be defined for the selected feature vector. This matrix constitutes the diffusion kernel. As proposed in [7, 8], the *kl*th element of matrix $\mathbf{W}$ is calculated according to:

$$\mathbf{W}_{kl} = \frac{\pi}{d_{kl}} \exp\left\{ -\frac{(\overline{\mathbf{T}}_k - \overline{\mathbf{T}}_l)^T [\hat{\boldsymbol{\Sigma}}_k + \hat{\boldsymbol{\Sigma}}_l]^{-1} (\overline{\mathbf{T}}_k - \overline{\mathbf{T}}_l)}{\varepsilon} \right\} \tag{5}$$

where ε is the kernel scaling factor that conveys a tradeoff between integration of large number of samples (large scale), and locality (small scale), and d_{kl} is the following normalization factor:

$$d_{kl} = \sqrt{\det\left(\mathrm{Cov}\left(\frac{\overline{\mathbf{T}}_k + \overline{\mathbf{T}}_l}{2}\right)\right)}. \tag{6}$$

Next $\{\lambda_j\}_{j=0}^{m-1}$ and $\{\varphi_j\}_{j=1}^{m-1}$ the eigenvalues and the eigenvectors of the affinity matrix W are calculated. Choosing the principal eigenvectors, as suggested in [9], provides the basis for the representation of the data in terms of its independent controlling parameters. In our constellation, these vectors correspond to the desired location coordinates of the source. In most cases, the dominant vectors are attached to the largest eigenvalues. We reasonably assume that this principle holds here as well.

3.3. Test Stage

Given a new set of M observations taken from new *unknown* locations, we construct an *asymmetric* affinity matrix between the feature vectors of the entire set (i.e. both the training and the test) and the training set:

$$\mathbf{A}_{kl} = \exp\left\{ -\frac{(\mathbf{T}_k - \overline{\mathbf{T}}_l)^T \hat{\mathbf{\Sigma}}_l^{-1} (\mathbf{T}_k - \overline{\mathbf{T}}_l)}{\varepsilon} \right\}. \qquad (7)$$

This asymmetric kernel is normalized to $\tilde{\mathbf{A}} = \mathbf{A}\mathbf{S}^{-\frac{1}{2}}$,where $\mathbf{S}$ is the diagonal matrix $\mathbf{S} = \text{diag}\{\mathbf{A}^T\mathbf{A}\mathbf{1}\}$ with $\mathbf{1}$ a vector of all '1's. The normalized matrix satisfies $\mathbf{W} = \tilde{\mathbf{A}}^T\tilde{\mathbf{A}}$. Therefore, both the right singular-vectors of $\tilde{\mathbf{A}}$ and the eigenvectors of $\mathbf{W}$ are identical. In addition, we can represent the left-singular vectors $\boldsymbol{\psi}_j$ of $\tilde{\mathbf{A}}$ as a weighted interpolation of the eigenvectors of $\mathbf{W}$, by:

$$\boldsymbol{\psi}_j = \frac{1}{\sqrt{\lambda_j}} \tilde{\mathbf{A}} \boldsymbol{\varphi}_j. \qquad (8)$$

Eventually, assisted by these relations, we reconstruct an *embedding* of the measurements onto the space spanned by the d left-singular vectors in correspondence with the dimensions of the parameter space:

$$\Psi : \mathbf{T}_i \mapsto \left[\boldsymbol{\psi}_1^{(i)}, \ldots, \boldsymbol{\psi}_d^{(i)}\right]^T. \qquad (9)$$

where $\boldsymbol{\psi}_1^{(i)}$ denotes the ith entry of the vector $\boldsymbol{\psi}_1$. In order to estimate the original location parameters, which are presumably responsible for generating the test data, we interpolate the training locations according to distances in the embedded space:

$$\hat{\boldsymbol{\theta}}_i = \sum_{\overline{\mathbf{T}}_j \in \mathcal{N}_i} \gamma_j(\mathbf{T}_i)\bar{\boldsymbol{\theta}}_j. \qquad (10)$$

where $\mathcal{N}_i$ consists of the k nearest training measurements $\overline{\mathbf{T}}_j$ of $\mathbf{T}_i$, in the embedded space. The interpolation coefficients are given by:

$$\gamma_j(\mathbf{T}_i) = \frac{\exp\left(-\|\Psi(\mathbf{T}_i) - \Psi(\overline{\mathbf{T}}_j)\|^2/\varepsilon_{\gamma_j}\right)}{\sum\limits_{\overline{\mathbf{T}}_s \in \mathcal{N}_i} \exp\left(-\|\Psi(\mathbf{T}_i) - \Psi(\overline{\mathbf{T}}_s)\|^2/\varepsilon_{\gamma_s}\right)} \qquad (11)$$

where ε_{γ_j} is the local variance in a neighborhood around $\Psi(\overline{\mathbf{T}}_j)$. Accordingly, the normalized estimation error is defined by:

$$e(\mathbf{T}_i) = \frac{\left\|\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i\right\|}{\|\boldsymbol{\theta}_i\|}. \qquad (12)$$

Hence, the root mean square error (RMSE) is given by:

$$\text{RMSE} = \sqrt{\frac{1}{M}\sum_{i=1}^{M} e^2(\mathbf{T}_i)}. \qquad (13)$$

3.4. Gaussian Mixture Model (GMM) Interpretation

The Gaussian kernel and the extension scheme can be interpreted as an implicit GMM. At the training stage, each training sample defines a local Gaussian surrounding it, i.e., $\mathcal{T}_l = \mathcal{N}(\overline{\mathbf{T}}_l, \hat{\mathbf{\Sigma}}_l)$. Accordingly, in the test stage, the Gaussian kernel based on the Mahalanobis distance in (7) represents the probability of each test sample $\mathbf{T}_k$ to be associated with the local Gaussian represented by the training sample $\overline{\mathbf{T}}_l$. Thus, we can redefine (7) as $A_{kl} = \Pr(\mathbf{T}_k | \mathbf{T}_k \sim \mathcal{T}_l)$.

4. EXPERIMENTAL RESULTS

In this section we examine the capability of the proposed algorithm to recover the location of an acoustic source, through a MATLAB simulation. The performance of the algorithm is inspected in both 2-D (azimuth and elevation) and 1-D (only azimuth) settings. In some scenarios, e.g. meeting rooms or cars, it can be assumed that the speaker location is confined to a predefined area.

4.1. Two-Dimensional Case

In the first experiment we demonstrate the applicability of the proposed algorithm to localize a source in a two-dimensional problem, where both the azimuth and the elevation angles are unknown. We test the ability of the presented method to organize the recordings according to these two angles. For generating the data, we use the image method [10][1]. Room dimensions are set to $[6, 5, 3.5]$ m, and the two microphones are located at $r_1 = [3, 3, 1]$ m and $r_2 = [3.2, 3, 1]$ m, respectively. The reverberation time of the room is set to $T_{60} = 0.3$ sec, emulating moderate reverberation conditions.

First, we generate $m = 481$ training samples located on a sector of a sphere around the first microphone. This sphere has a fixed radius of 1 m, while the azimuth angle ranges between $[\pi/16, 2\pi/16]$ rad and the elevation angle ranges between $[7\pi/16, \pi/2]$ rad. We generate the AIRs $\bar{h}_{\theta,ki}(n)$ in accordance with each parameter combination. As an input in the training stage we generate an arbitrary WGN 3 sec long signal with sampling rate 8000 Hz and convolve it with $\bar{h}_{\theta,ki}(n)$ to obtain the measured signals $\bar{y}_{ki}(n)$. For each location, we create an additional $L = 20$ measurements with low variance Gaussian location perturbations. Next, we simulate another set of $M = 480$ samples from other locations in the designated range, constituting the test set for examining the performance of the proposed algorithm. In order to generate the test set, we play from each location a *different* 3 sec long speech signal. Each signal is convolved with $h_{\theta,ki}(n)$, resulting in the output signals $y_{ki}(n)$ as received by the microphones. For each measurement we estimate the RTF according to Welch's averaged periodogram method with an Hamming window of length 256, 50% overlap between adjacent sections, and $D = 256$ frequency bins.

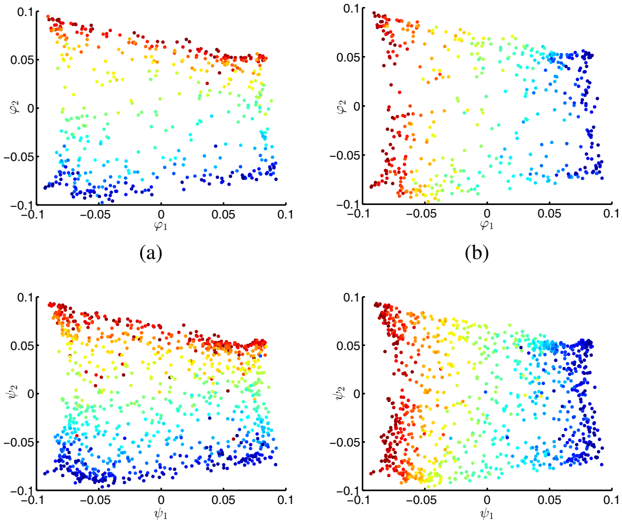
The embedding of the parameters of only the training set and the entire set, is shown in Fig. 1. The coloring of the points according to the true values of the angles indicates that the embedding captures the independent controlling parameters, as the coloring scheme is maintained. Re-parameterization of the test set yields a mean error of RMSE $= 0.0079$ rad.

Note that conventional localization methods that use two microphones are incapable of localizing a source in the two-dimensional case, but merely the TDOA value.

4.2. One-Dimensional Case

In the second experiment we check the ability to recover the azimuth angle, while the elevation angle is fixed to $\pi/2$ rad. The results of the proposed algorithm are compared with the classical generalized cross-correlation (GCC)-maximum likelihood (ML) algorithm [1], for different reverberation times. The room setup is slightly changed to make the reverberation influence more noticeable. Accordingly, the distance between the speaker and the mi-

[1] We used an efficient implementation provided by E.A.P. Habets at http://home.tiscali.nl/ehabets/rir_generator.html



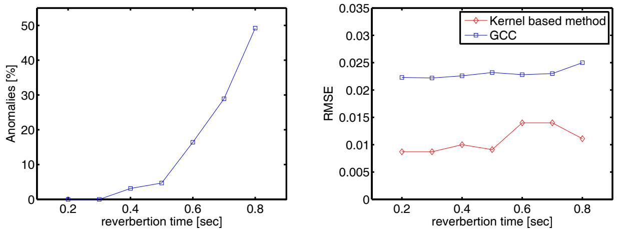
(a)

(b)

(c)

(d)

Figure 1: Scatter plot of the embedding of the training and test sets. Training samples only in (a),(b) and the entire training and test sets in (c),(d). Color coding according to the values of the elevation angle in (a),(c) and the azimuth angle in (b),(d).



(a) Anomalies of the GCC.

(b) RMSE of GCC and kernel-based.

Figure 2: The Anomaly percentage and the RMSE as a function of the reverberation time for GCC and the kernel-based algorithms.

crophones is enlarged by 1 m, and the room dimensions are set to $[6, 5, 3]$ m. For each reverberation time, we simulate 256 samples in the range $[\pi/16, 4\pi/16]$ rad, out of which 50% are treated as a test set, while the rest defines the training set. The sampling rate used for the GCC is set to 44100 Hz, however, for the kernel-based algorithm the rate is maintained at 8000 Hz to reduce the computational complexity. The results of both algorithms, are shown in Fig. 2. In Fig. 2(a), we show the anomaly percentage, defined as the percentage of experiments for which the algorithms exceed a pre-defined error threshold, set to $\mathbf{10\%}$. No anomaly was detected for the proposed algorithm. In Fig. 2(b), the RMSE of estimates within the anomaly threshold is presented. It is clearly seen that the proposed algorithm outperforms the GCC method in both figures-of-merit. In high reverberation, the GCC is incapable of distinguishing between the direct arrival and the reflections. A misidentification of the direct path, may result in a large estimation error. The proposed algorithm is more robust to reverberation, since the variations in the entire RTF are taken in account.

5. CONCLUSIONS

A novel approach for the well-studied problem of source localization was presented. The proposed approach utilizes state-of-the-art manifold learning techniques and extends previous results to work with speech signals by utilizing two microphones. In the proposed algorithm, we first define a feature vector comprised of spectral bins of the RTF relating the two microphones. This feature embodies the independent parameters that control our system and cancel out the effect of the varying PSD of the speech signal. Given this feature vector, we construct a diffusion kernel and show that its eigenvectors form the basis for spanning the location coordinates in the parametric space. In fact, the training data, which are given beforehand, have a major role in establishing meaningful relations between distances in the parametric space and in the observable space. Experimental results demonstrate the efficacy of the algorithm, especially in reverberant environments, compared with the classical GCC method. Moreover, the algorithm can reveal both azimuth and elevation angles using merely a microphone pair, an impossible task with traditional methods. The encouraging results might pave the road to solving more realistic scenarios in which the training and test conditions are acoustically different.

6. REFERENCES

[1] C. Knapp and G. Carter, “The generalized correlation method for estimation of time delay,” *IEEE Trans. on Acoust., Speech, Signal Process.*, vol. 24, no. 4, pp. 320–327, Aug. 1976.

[2] J. Chen, J. Benesty, and Y. Huang, “Time delay estimation in room acoustic environments: an overview,” *EURASIP J. Appl. Signal Process.*, vol. 2006, pp. 170–170, Jan. 2006.

[3] T. Dvorkind and S. Gannot, “Time difference of arrival estimation of speech source in a noisy and reverberant environment,” *Signal Processing*, vol. 85, no. 1, pp. 177–204, Jan. 2005.

[4] R. Talmon, D. Kushnir, R. Coifman, I. Cohen, and S. Gannot, “Parametrization of linear systems using diffusion kernels,” *IEEE Transactions on Signal Processing*, vol. 60, no. 3, pp. 1159–1173, Mar. 2012.

[5] R. Talmon, I. Cohen, and S. Gannot, “Supervised source localization using diffusion kernels,” in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2011, pp. 245–248.

[6] A. Deleforge and R. Horaud, “2D sound-source localization on the binaural manifold,” in *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2012, pp. 1–6.

[7] A. Singer and R. Coifman, “Non-linear independent component analysis with diffusion maps,” *Appl. Comput. Harmon. Anal.*, vol. 25, pp. 226–239, 2008.

[8] D. Kushnir, A. Haddad, and R. R. Coifman, “Anisotropic diffusion on sub-manifolds with application to earth structure classification,” *Applied and Computational Harmonic Analysis*, vol. 32, no. 2, pp. 280–294, 2012.

[9] A. Singer, “Spectral independent component analysis,” *Appl. Comput. Harmon. Anal.*, vol. 21, pp. 128–134, 2006.

[10] J. Allen and D. Berkley, “Image method for efficiently simulating small-room acoustics,” *J. Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, Apr. 1979.