Applied Acoustics 197 (2022) 108918

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Single-sensor localization of moving acoustic sources using diffusion kernels

Eran Zeitouni *, Israel Cohen

Andrew and Erna Viterbi Faculty of Electrical & Computer Engineering, Technion–Israel Institute of Technology, Haifa 3200003, Israel

ARTICLE INFO

Article history:
Received 17 October 2021
Received in revised form 9 June 2022
Accepted 2 July 2022

Keywords:
Source localization
direction finding
single-sensor
Single-site
Manifold learning
Diffusion maps
Passive sensing
Position finding
Non-cooperative localization

ABSTRACT

Source localization is a common problem in various fields and has applications in both military and civil sectors. Localization of acoustic sources generally requires a few microphones, but it is also possible to use a single microphone and data that was prerecorded in the same environment. Unfortunately, existing single-microphone localization methods are restricted to acoustic sources that have a fixed location. In this paper, we introduce a supervised method for estimating both the location and velocity of a moving acoustic source, using a single microphone based on a manifold learning approach. Simulation results demonstrate the sensitivity of the algorithm to variations in the speed of the sources, resulting in a trade-off between the accuracy of the estimated location and the accuracy of the estimated direction. In addition, the results demonstrate the sensitivity to variations in direction and frame length of the received signal. The algorithm performs well in reverberant and noisy environments, yet is sensitive to environmental conditions changes.

1. Introduction

Source localization problem has drawn attention and extensive research efforts over the last several decades. Its applications are tremendously broad and vary generally by the type of source signal, e.g., radio frequency or acoustic, and by the goals of the interested parties. Solutions to the source localization problem can be divided into two groups: active sensing and passive sensing. Classic and modern passive source localization methodologies mostly rely on exploiting variations of a single physical attribute between the received signals, which originated from a particular source. Among these attributes are the amplitude (e.g., Watson-Watt technique), frequency (e.g., Doppler effect), phase (e.g., correlative interferometer method), and time (e.g., time difference of arrival, TDOA). Except for the latter, utilization of these variations by a sensor enables to estimate the direction toward the signal source. This process is commonly known as direction finding (DF), where the determined direction toward the source is called either bearing or direction of arrival (DOA). Consequently, the source location is determined as the point of intersection of bearings produced by a spatial array of sensors that form triangulation. Theoretically, the point of intersection (i.e., a fix) is determined by at least three bearings. But in practice, under sensors' formations that are compatible with reasonable operational scenarios, two bearings are mostly sufficient. On the other hand, utilization of the acquisition time difference between the sensors results in a hyperbola. Accordingly, in order to determine the location of the source, at least three hyperbolas are needed for an unambiguous point of intersection. Thus, source localization by these passive methodologies is constrained to multiple sensors: Either two sensors (each composed of a spatial array of multiple inner elements, such as antennas or microphones) for all cases except TDOA, or three sensors for TDOA (where each sensor involves a single element).

In recent years, advanced spatial array processing methods and algorithms for localization have been introduced, such as the maximum likelihood (ML) based beamforming and subspace-based methods. The former relies on optimizing the output power of the beamformer, according to a statistical model of the received signals. A popular algorithm that follows this approach is the steered response phase transformation (SRP-PHAT) [1,2]. On the other hand, the subspace based approaches, such as the well-known multiple signal classification (MUSIC) algorithm [3], successfully tackle ML-based algorithms' fundamental flaws, alongside yielding high-resolution results. These benefits of the subspace-based methods are obtained even in the presence of

* Corresponding author.
E-mail addresses: eranze@campus.technion.ac.il (E. Zeitouni), icohen@ee.technion.ac.il (I. Cohen).

https://doi.org/10.1016/j.apacoust.2022.108918
0003-682X/

noise, but they come at the price of high computational and storage resources. Common to all these algorithms is the requirement of a spatial array.

In contrast to the mostly well-posed source localization problem of a pair of sensors (where each is composed of a spatial array), under the constraint of a single sensor (e.g., a car speakerphone, operational scenarios where only a single sensor is available) this problem is mostly ill-posed/underdetermined using conventional localization methods. The reason for that is that the system transfer function (infamous multipath included) is mostly unknown. In case it is known, alongside the statistical model of the transmitted signal, the location can be recovered by ML-based beamformer.

Another approach [4–6], [7, Ch. 7], [8, Ch. 7] for source localization using a single sensor (composed of a spatial array of multiple antennas) is restricted to the high frequency (HF) range, due to the modes of propagation of the signals in this range. The main propagation component in the HF range, skywave, is reflected back from the ionosphere toward Earth's surface. Thus, in addition to measuring the DOA, by recovering the elevation angle, a triangulation is achieved and the location can be estimated. However, in order to do so, the virtual height of the reflecting ionosphere layer must be known and since it depends on many variables (e.g., weather and solar activity), it is mostly impractical. In addition to the previously mentioned single-sensor source localization methods, a useful technique named "running fix" [8, Ch. 7], can be applied to a stationary (or relatively slow) source, using a moving sensor, if the source is active long enough.

Talmon et al. [9] have introduced a supervised manifold learning-based method, which was implemented explicitly by diffusion kernels. Instead of fitting the identified system to a predefined model as conventional system identification methods, this data-driven method focuses on revealing the underlying fundamental controlling parameters of the system. To do so, this method is based on a training set of measurements of signals, such that no knowledge regarding system transfer function is necessary. The diffusion kernel, which is based on a specially-tailored distance measure, combines local estimates of covariance matrices of the measurements with global processing by spectral decomposition. Accordingly, the diffusion kernel allows parameterization of the measurements into a low-dimensional space. This low-dimensional space is also known as the manifold. The unknown parameter can be estimated based on the relation between its embedded representation to the corresponding representations associated with the training set. This method was later applied by Talmon et al. [10] for source localization of stationary (i.e., fixed location) acoustic sources by a single microphone, in a small reverberant room. Since the azimuth angle of the location of the sources is the only degree of freedom of the system, the unknown location can be estimated using this method. This method was further investigated in a series of works by Laufer-Goldshtein et al. for source localization of stationary acoustic sources by a pair of sensors [11–14], and moving sources using multiple pairs of sensors [15–17]. All of Laufer-Goldshtein's works were recently concluded in [18].

Another subspace-based approach with a similar concept, called fingerprinting, is common in the localization of RF emitters in highly scattering environments. As its name implies, this technique assumes a unique relation between the emitter location and the characteristics of the signals intercepted by the sensor (mostly multipath propagation). First, a database of fingerprints, extracted from signals received from a specific area, is composed in advance. Subsequently, the unknown location can be recovered by matching its corresponding fingerprint to the fingerprints in the database. Several works have focused on localization using a single sensor [19–26] in recent years- all composed of a spatial array of multiple antennas. However, it is worth mentioning that one work, by Kupershtein et al. [19] has tried to determine the location using a single element as well- mostly with no success, in addition to significant performance degradation compared to multiple elements.

A different single-sensor localization approach determines the location using aids. As standard spatial array-based methods, it provides a DOA according to the direct propagation path of the signal transmitted by the emitter. In addition, this technique exploits either a known dominant scatterer [27] or a set of known transponders [28,29] available at the scene, in order to produce a fix.

In this paper, the paradigm of [10] is adopted and adapted to a more realistic setting, which corresponds to scenarios of moving acoustic sources in a small reverberant environment, while the sensor remains stationary (for simplicity). Thus, the new setting has led to three major changes in the algorithm of [10]; 1) In the case of sources that have a fixed location, an additional varying degree of freedom of the system is assigned for the distance of the source from the sensor. The addition of a second degree-of-freedom assembles the two-dimensional location of the source, along with the existing azimuth degree-of-freedom. On the other hand, in the case of moving sources, the sole degree-of-freedom is allocated for one of the velocity's parameters: either speed or direction of movement. This velocity parameter dictates the location parameters of the sources throughout their movement. 2) The system is time variant, by definition, due to the movement of the sources. It prevents the use of familiar and convenient convolution-based relations. 3) The Mahalanobis distance-based diffusion kernel is replaced by a customized Euclidean distance-based diffusion kernel. Our kernel is inspired by the study of the acoustic manifold by Laufer-Goldsthtein et al. [12] and its successful application in their following paper [13]- both focusing on multiple sensors. However, the extension of the diffusion framework for the test set, and its corresponding diffusion kernel, in particular, do not follow either [10] or [12,13]. The results of our algorithm in a simulated reverberant environment demonstrate accurate single-sensor source localization for moving sources, under various conditions. For clarity, we note that in the rest of this paper, we refer to a sensor and a microphone interchangeably.

The outline of this paper is as follows. In Section 2, we formulate the problem. In Section 3, the computation of the diffusion kernel is presented. In Section 4, we present the proposed algorithm for single-sensor source localization for moving sources. In Section 5, results of the proposed algorithm are shown and compared. In Section 6, conclusions regarding the performance are presented.

2. Problem formulation

Throughout this paper, matrices are denoted by bold capital letters, whereas vectors are denoted by bold small letters. Moreover, elements in matrices and vectors are written with a superscript index in parentheses, e.g., the ith element of the vector $\mathbf{a}$ is expressed as $\mathbf{a}^{(i)}$. We consider a standard enclosure, such as a conference room. Each source, one at a time, transmits a signal during its movement. We assume that the source signal is a zero-mean wide-sense stationary (WSS) process. Since many natural signals, such as speech and music, are WSS in short time frames (i.e., quasi-stationary process), we can even be satisfied with such a weaker assumption.

An acoustic impulse response (AIR), between a source and the microphone, is affected by several factors, such as environment dimensions, locations of the source and the microphone, reflection coefficients (or reverberation time) of the walls, floor, and ceiling, and the presence of objects in the room. Let $h_{\theta_i(j)}(n,j)$ denote a real-valued AIR, which is defined as the response at discrete time index n to an impulse transmitted at discrete time index j, between

the ith source and the microphone, with respect to the parameters vector $\boldsymbol{\theta}_i(j)$. The parameters vector of the ith source at discrete time index j, is defined as a combination of the relative location and the velocity of the source, i.e., $\boldsymbol{\theta}_i(j) = [\rho_i(j), \phi_i(j), s_i(j), \beta_i(j)]$, where $\rho_i(j)$ is the distance (i.e., radius) between the source and the microphone; $\phi_i(j)$ is the bearing angle (also azimuth in our setting); $s_i(j)$ is the speed of the source; and $\beta_i(j)$ is the direction of movement (also known as the course) of the source. It is assumed that the height difference between the source and the microphone is negligible- thus the elevation angle remains constant. Note that $\boldsymbol{\theta}_i(j)$ can also be represented by Cartesian coordinates.

The signal received by the sensor, denoted by $y_i(n)$, consists of the direct and indirect propagation paths of the transmitted signal, and is defined by

$$y_i(n) = H_{\boldsymbol{\theta}}\{x_i(n)\} = \sum_{j=-\infty}^{\infty} h_{\boldsymbol{\theta}_i(j)}(n,j)x_i(j), \tag{1}$$

where $x_i(n)$ is the signal transmitted by the ith source. We note that $x_i(n)$ and $y_i(n)$ are the real-valued input and output signals, of finite length, of the system $H_{\boldsymbol{\theta}}$, which corresponds to the AIR and depends on the parameters vector. In addition, since the AIR is affected tremendously by the movement of the source, the system is time variant. We use a white Gaussian noise (WGN) signal as the source signal, since it fully excites the frequency response of the AIR. The received signal is saved and divided into time frames.

We assume the trajectory, formed by the movement of the source during the time frame, can be approximated by a linear movement segment. We inspect each time frame individually for estimation of the location and velocity. The goal of the proposed algorithm is to determine the unknown location and velocity of a source based on a training dataset, which is available beforehand. For each time frame, we manage two datasets, based on the different signals received by the sensor: a training dataset and a test dataset. In order to generate the training dataset, we choose arbitrarily m known locations and velocities of the source $\overline{\boldsymbol{\Theta}} = \{\overline{\boldsymbol{\theta}}_1(q), \ldots, \overline{\boldsymbol{\theta}}_m(q)\} \subset \mathbb{R}^d$, where q is a query point along the trajectories of the sources, and d is the dimension of the parameters vector (i.e., the number of system's degrees of freedom).

Let $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_{m+1}(q), \ldots, \boldsymbol{\theta}_{m+M}(q)\} \subset \mathbb{R}^d$ be a set of M arbitrary unknown source locations and velocities, corresponding to the M measurements of the test dataset. We define the query point q as the midway point of the trajectory, such that the approximation error of the true trajectory by the linear segment is minimized. Note that the acoustic environment is fixed between training and test stages (i.e., room characteristics and microphone location remain unchanged), thus the only degrees of freedom of the controlling parameters of the AIR are the locations of the sources and their velocities.

3. Diffusion kernel

In this section, we present a diffusion kernel between feature vectors, derived from the given observations.

3.1. From observations to feature vectors

We follow Talmon et al. [9,10], and define our feature vector based on an autocorrelation function of the observation. The reason for that choice is that a second-order statistics measure conveys the location better and is less dependent on the specific random unknown transmitted signal, rather than using the raw observation. From (1), under the assumption of a WGN input signal, the autocorrelation function of $y_i(n)$, the output signal of a time-variant system, is given by

$$\begin{aligned} c_{\mathbf{y}_i}(n_1,n_2) = \ & \mathbb{E}[y_i(n_1)y_i(n_2)] == \sum_{j,l=-\infty}^{\infty} h_{\boldsymbol{\theta}_i(j)}(n_1,j)h_{\boldsymbol{\theta}_i(l)}(n_2,l)c_{\mathbf{x}_i}(j-l) = \\ & = \sigma^2_{\mathbf{x}_i}\sum_{j=-\infty}^{\infty} h_{\boldsymbol{\theta}_i(j)}(n_1,j)h_{\boldsymbol{\theta}_i(j)}(n_2,j), \end{aligned} \tag{2}$$

where $c_{\mathbf{x}_i}(j-l)$ and $c_{\mathbf{y}_i}(n_1,n_2)$ denote the time-invariant autocorrelation function of the input signal $x_i(n)$ ($c_{\mathbf{x}_i}(\tau) = \sigma^2_{\mathbf{x}_i}\delta(\tau)$ for WGN) and the time-variant autocorrelation function of the output signal $y_i(n)$, respectively. $\mathbb{E}[\cdot]$ denotes an expected value. As implied by (2), we can represent the observation $\mathbf{y}_i$ as a function of the controlling parameters $\boldsymbol{\theta}_i$ of the system. It is assumed that given a sufficiently short time interval, the first two moments (autocorrelation in particular) of the quasi-stationary input signal would not change along the interval.

In fact, by considering an additional assumption (apart from short time intervals) of slow speed and gradually changing velocity, we introduce small changes to the AIR along the time frames. As a result, we can obtain the familiar convolution-based version of (2), as in [9,10]:

$$c_{\mathbf{y}_i}(\tau) = h_{\boldsymbol{\theta}_i}(\tau) * h_{\boldsymbol{\theta}_i}(-\tau) * c_{\mathbf{x}_i}(\tau). \tag{3}$$

As indicated by (3), the time differences of the autocorrelation function of the output signal depend purely on the variations of the AIR, i.e., on the evolution of the location and velocity of the ith source, $\boldsymbol{\theta}_i$.

Let c express the nonlinear mapping of the location and the velocity of the ith source, $\boldsymbol{\theta}_i \in \mathbb{R}^d$, to the first D elements of the autocorrelation function of the observation $\mathbf{y}_i$, defined as

$$\mathbf{c}_i = c(\boldsymbol{\theta}_i), \tag{4}$$

where $\mathbf{c}_i \in \mathbb{R}^D$ is a vector of length D which composed of the autocorrelation elements, i.e.,

$$\mathbf{c}_i^{(j)} = c_{\mathbf{y}_i}(n, n+j) = \mathbb{E}[y_i(n)y_i(n+j)] \tag{5}$$

for $j = 0, \ldots, D-1$. In such a manner, we extract a feature vector for each signal received by the sensor. Note that the length of the feature vectors, D, should reflect the tradeoff between the length of the autocorrelation function (large value) and the latency and quasi-stationarity properties considerations (small value). Let $\Gamma = \{\mathbf{c}_i\}_{i=1}^M$ be the set of the feature vectors with respect to the unlabeled parameters in $\boldsymbol{\Theta}$. Accordingly, let $\overline{\Gamma} = \{\overline{\mathbf{c}}_i\}_{i=1}^m$ denote the set of the feature vectors with respect to the labeled parameters in $\overline{\boldsymbol{\Theta}}$. We aim to recover the unknown parameters vectors based on the aforementioned feature vectors.

3.2. Manifold structure and the choice of an affinity measure

As specified in Sections 3.1 and 2, the autocorrelation function based feature vectors have a high-dimensional representation in $\mathbb{R}^D$, which resembles the high number of reflections from all surfaces characterizing the bounded environment- hence the AIR. On the other hand, the typical AIR, associated with our feature vector, is characterized by an exponentially decaying envelope. Moreover, the feature vectors are affected by a small set of parameters associated with the physical attributes of the enclosure, in addition to the influence of the speech signal of the source which is quasi-stationary. Therefore, it is assumed that not only that the feature vectors, which were originated from a specific region of interest in the enclosure, do not spread uniformly in the entire space of $\mathbb{R}^D$, but are also restricted to a more explicit and even compact structure. This structure, namely the manifold $\mathscr{M}$ of dimension d,

is significantly smaller than the dimension of the surrounding high-dimensional space (i.e., $d \ll D$). Thus, by applying the notation from Section 3.1, we define $c : \Theta \to \Gamma$ to be the nonlinear map between an unknown parametric manifold $\mathscr{M} \subseteq \boldsymbol{\Theta} \subset \mathbb{R}^d$ and its corresponding observation-based feature vectors dataset $\Gamma \subset \mathbb{R}^D$.

Even though our setting involves moving sources, the hypothesis of such manifold can be rationalized by the combination of quasi-stationary input signal, slow speed and gradually changing velocity components of the source, stationary microphone and short time intervals. These assumptions allow us to experience small changes in the feature vector along the time intervals, and thus inferring the locations and velocities of the sources, which are the only degrees of freedom of the system. Thus, we conclude that the feature vectors can be represented by a low-dimensional manifold, whose embeddings are ruled by the location and velocity of the sources.

The considered low-dimensional manifold is assumed to be a real nonlinear structure, but in practice, it is locally linear in small areas. Indeed, the surface of the manifold is flat in the close neighborhood of each embedded point and coincide with the tangent plane to the manifold at this embedded point. This is in line with the implied assumption, based on which a small change in the physical parameters vector leads to a slight change in the corresponding feature vector, whereas a big variation results in a completely different feature vector. Therefore, the similarities between points that are located on the manifold in the vicinity of each other, can be measured reliably by using the Euclidean distance. Note the Euclidean distance cannot assess reliably affinity for large scales. Instead, it shall be dealt with the geodesic distance, which is the locally shortest path along the manifold, in the case that the structure of the manifold is known (the Euclidean distance is equal to the geodesic distance in the case of a flat/linear manifold only).

Another popular choice for measuring affinities on the manifold is Mahalanobis distance, which was the key element in several papers [9–11], since the affinities measured between feature vectors by this distance approximate the Euclidean distance between the corresponding physical parameters vectors [30]. However, it holds several fundamental practical drawbacks: In order to estimate its local covariance matrices, several additional local slightly perturbed observations shall be generated for each training observation, which is in fact a resources-related burden in terms of both storage and implementation during a real-data experiment. In addition, these local covariance matrices are singular and thus not invertible, whereas computing their pseudo-inverse (their rank is d [9]) as an alternative is not necessarily trustworthy. In this work, for complying with the nonlinear and unknown structure of the manifold, affinities in local vicinities are measured using Euclidean distance, whereas greater distances are omitted.

3.3. Diffusion kernel computation

In order to acquire the independent parameters controlling our system, which their availability is solely by the nonlinear feature vectors of the observations, we define an $m \times m$ affinity matrix $\mathbf{W}$ between all the feature vectors in $\overline{\Gamma}$, related to the corresponding parameters vectors set $\overline{\boldsymbol{\Theta}}$. The affinity matrix consists of a Gaussian kernel with a scale parameter ε, and following Section 3.2 its ijth element is given by

$$\mathbf{W}^{(ij)} = \begin{cases} \exp\left\{-\frac{||\bar{\mathbf{c}}_i - \bar{\mathbf{c}}_j||^2}{\varepsilon}\right\}, & \text{if } \bar{\mathbf{c}}_i \in \overline{\mathscr{N}}_j \text{ or } \bar{\mathbf{c}}_j \in \overline{\mathscr{N}}_i \\ 0, & \text{otherwise,} \end{cases} \tag{6}$$

where $\overline{\mathscr{N}}_j$ is the set of the k-nearest-neighbors of $\bar{\mathbf{c}}_j$ in $\overline{\Gamma}$. It is worth mentioning that the set $\overline{\mathscr{N}}_j$ is selected by the ordinary Euclidean distance between the feature vectors, since as mentioned previously it is trustworthy in close neighborhoods only.

The value of ε is determined in proportion to the median value of Gaussian function's various numerator values associated with the non-zero elements of the affinity matrix. That is reasoned by the assumption that the Euclidean distance is monotonic with respect to the parameters vector in small areas, as discussed in Section 3.2. The proportion is decided by an exhaustive search.

4. Localization based on diffusion mapping

In this section, we introduce a supervised single-sensor source localization algorithm for recovering the location and velocity of the sources by utilization of the feature vectors, based on the kernel of Section 3. Thanks to the eigendecomposition of the kernel, we find the mapping from the observable space $\mathbb{R}^D$ to the intrinsic manifold space of $\mathbb{R}^d$, which is ruled by the dominating parameters vector and resembles the original parametric space up to a monotonic distortion. Based on the mapping of all training observations into the manifold, the localization of a new observation of unknown parameters vector is estimated by exploiting its nearest labeled neighbors in the embedded space.

4.1. Manifold parameterization

By normalizing the affinity matrix $\mathbf{W}$, using a diagonal matrix $\mathbf{D}$ with $\mathbf{D}^{(ii)} = \sum_{j=1}^{m} \mathbf{W}^{(ij)}$, we obtain the transition matrix

$$\mathbf{P} = \mathbf{D}^{-1}\mathbf{W}, \tag{7}$$

which defines a Markov process, namely a discrete diffusion process over the training dataset. Using the transition matrix $\mathbf{P}$, we establish the normalized graph-Laplacian $\mathbf{L}$ [31], by $\mathbf{L} = \mathbf{I} - \mathbf{P}$, where $\mathbf{I}$ is an identity matrix. It can be shown, under certain conditions, that the graph-Laplacian matrix $\mathbf{L}$ converges to the Fokker–Planck operator on the manifold [32,33], which describes a continuous diffusion process over the dataset.

By applying eigendecomposition of the transition matrix $\mathbf{P}$, the labeled feature vectors are nonlinearly mapped into a new embedded space, according to the parameterization of the manifold $\mathscr{M}$. In fact, the parameterization of the manifold constitutes an intrinsic representation of the labeled feature vectors. Let $\{\lambda_j\}_{j=0}^{m-1}$ and $\{\boldsymbol{\psi}_j\}_{j=0}^{m-1}$ be the eigenvalues and eigenvectors of the transition matrix $\mathbf{P}$. Note that $\lambda_0 = 1$ and its corresponding eigenvector $\boldsymbol{\psi}_0$ is a vector of ones [34]. The eigenvectors of $\mathbf{P}$ are assumed to establish the reparameterization of the independent controlling parameters of the m training observations. Thus, let Ψ_d be the diffusion mapping of the training feature vectors into the embedded Euclidean space $\mathbb{R}^d$, which is spanned by d eigenvectors corresponding to the d largest eigenvalues (trivial case excluded). Ψ_d is defined as

$$\Psi_d : \bar{\mathbf{c}}_i \to \left[\lambda_1 \boldsymbol{\psi}_1^{(i)}, \ldots, \lambda_d \boldsymbol{\psi}_d^{(i)}\right]^T. \tag{8}$$

In fact, this map can be regarded as an approximation of the inverse-map of the nonlinear function c, in addition to a monotonic distortion between the two parametric spaces (i.e., the original and the embedded manifold) [9, Fig. 1]. Note that the diffusion maps combines local relations by the construction of the affinity kernel with global processing by the spectral decomposition.

4.2. Extension for new observations

Given an additional set of M new sequential observations, generated from unknown locations and velocities of the sources, we seek to embed them as well in the low-dimensional manifold.

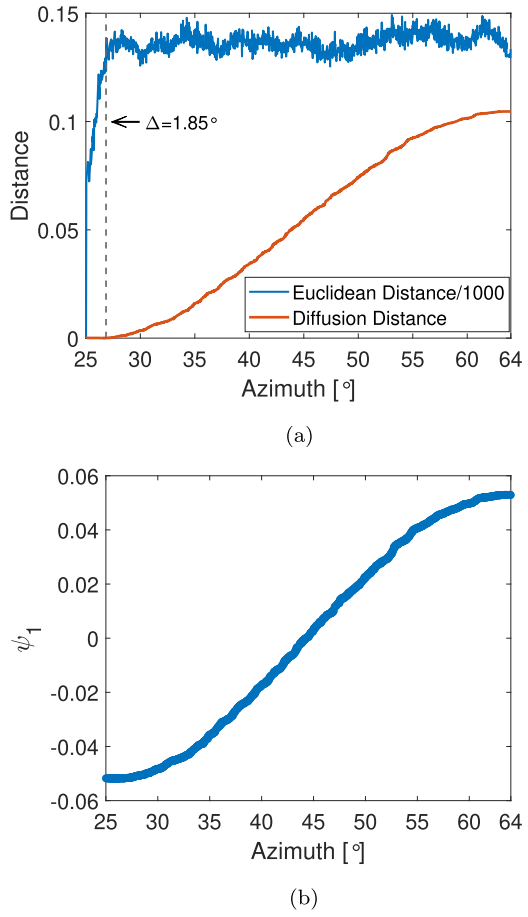

Fig. 1. (a) Comparison of the Euclidean (blue) and the diffusion (red) distances according to the azimuth angle. The distances are measured among all feature vectors with respect to a reference feature vector, representing the minimal angle (25°). The dashed line represents the range of monotonicity maintained by the Euclidean distance. (b) Embedding of the observations as a function of the azimuth angle.

However, in order to avoid another spectral decomposition, we tackle it according to Nyström method [35] for out-of-sample extension (OOSE), by adding M new rows to the affinity matrix $\mathbf{W}$:

$$\mathbf{W}^{(\tilde{i}j)} = \begin{cases} \exp\left\{-\frac{||\mathbf{c}_{\tilde{i}}-\bar{\mathbf{c}}_j||^2}{\varepsilon}\right\}, & \text{if } \bar{\mathbf{c}}_j \in \mathcal{N}_{\tilde{i}} \\ 0, & \text{otherwise} \end{cases}, \tag{9}$$

where $\tilde{i} = m + i$ and $\mathcal{N}_{\tilde{i}}$ is the set of the k-nearest-neighbors of $\mathbf{c}_i$ in $\overline{\Gamma}$. In contrast to the construction process of the symmetric affinity matrix $\mathbf{W}$ of (6), the affinity between the observations in (9) is implemented with respect to the training set only. Accordingly, the new entries of the transition matrix $\mathbf{P}$ are given by

$$\mathbf{P}^{(\tilde{i}j)} = \left(\sum_{j=1}^{m}\mathbf{W}^{(\tilde{i}j)}\right)^{-1}\mathbf{W}^{(\tilde{i}j)}. \tag{10}$$

The new entries of the extended eigenvectors of $\mathbf{P}$ can be represented as a weighted linear interpolation of the original entries of the eigenvectors:

$$\boldsymbol{\psi}_l^{(\tilde{i})} = \frac{1}{\lambda_l}\sum_{j=1}^{m}\mathbf{P}^{(\tilde{i}j)}\boldsymbol{\psi}_l^{(j)}. \tag{11}$$

Subsequently, Ψ_d maps the new feature vectors of the unlabeled observations to their corresponding representation of independent dominating parameters in the embedded manifold:

$$\Psi_d : \mathbf{c}_i \to \left[\lambda_1\boldsymbol{\psi}_1^{(\tilde{i})}, \ldots, \lambda_d\boldsymbol{\psi}_d^{(\tilde{i})}\right]^T. \tag{12}$$

The suggested extensions of former works were based on either additional mathematical relations [30,9–11] in the case of Mahalanobis distance-based diffusion kernel, or straightforwardly [13,18] in the case of Euclidean distance-based diffusion kernel. Note that our extended entries of the eigenvectors are obtained following normalization by the transition matrix. In addition, in order to maintain the nonlinear structure of the manifold, our extended entries of the eigenvectors are acquired by restricting the computation of the entries of the new rows of the affinity matrix to those which are associated with the nearest training observations. In other words, our extension is free from the issues of its Euclidean distance-based counterpart suggested in [13,18], whose combination of no normalization and affinity calculation with respect to all training observations (i.e., not restricted to close area only, in contrast to the gist of Section 3.2, as the Euclidean distance is accurate for larger distances in case of flat/linear manifold only) leads to a scaled and inaccurate extended manifold structure.

4.3. Recovery of the controlling independent parameters

We can take advantage of the proximity of the unlabeled test observations to the labeled training observations in the embedded manifold space $\mathbb{R}^d$, which were mapped by Ψ_d, for estimating their parameters vectors.

As mentioned in Section 3.2 regarding the structure of the manifold, the geodesic distance, which is the locally shortest path along the manifold, should be worked with in order to accurately measure affinities between feature vectors along the manifold. The geodesic distance can be approximated by the diffusion distance, which is equal to the Euclidean distance in the embedded space when using all eigenvectors. The diffusion distance can be appropriately approximated by using merely the first d non-trivial eigenvectors [36], i.e.,

$$D_{\mathrm{DIFF}}(\mathbf{c}_i, \mathbf{c}_j) \cong \left\|\Psi_d(\mathbf{c}_i) - \Psi_d(\mathbf{c}_j)\right\|, \tag{13}$$

where $\Psi_d(\mathbf{c}_i) = \left[\lambda_1\boldsymbol{\psi}_1^{(\tilde{i})}, \ldots, \lambda_d\boldsymbol{\psi}_d^{(\tilde{i})}\right]^T$.

The ability to measure distances along the manifold thanks to the diffusion distance, provides us the option to determine the affinities between the feature vectors correctly. Thus, samples which are close to each other on the low-dimensional manifold are expected to be acquired from physically adjacent locations and to hold similar velocities. Accordingly, the unknown parameters vector of the test observation can be estimated using its labeled neighbors on the manifold, by a weighted interpolation of their parameters vectors:

$$\hat{\boldsymbol{\theta}}_i(q) = \sum_{j:\,\Psi_d(\bar{\mathbf{c}}_j)\in\widetilde{\mathcal{N}}_i}\gamma_j(\mathbf{c}_i)\bar{\boldsymbol{\theta}}_j(q), \tag{14}$$

where $\widetilde{\mathcal{N}}_i$ consists of the $\tilde{k}$-nearest embedded training measurements $\{\Psi_d(\bar{\mathbf{c}}_j)\}_{j=1}^{\tilde{k}}$ of $\Psi_d(\mathbf{c}_i)$ according to the diffusion distance. The interpolation coefficients $\{\gamma_j\}_{j=1}^{\tilde{k}}$, satisfying $\sum_{j=1}^{\tilde{k}}\gamma_j(\mathbf{c}_i) = 1$, are proportional to the distance between $\Psi_d(\mathbf{c}_i)$ and its neighbors (i.e., the diffusion distance between the test observation and each of its labeled neighbors):

$$\gamma_j(\mathbf{c}_i) = \frac{\exp\left(-\frac{\left\|\Psi_d(\mathbf{c}_i)-\Psi_d(\bar{\mathbf{c}}_j)\right\|^2}{\varepsilon_{\gamma_i}}\right)}{\sum_{l:\,\Psi_d(\bar{\mathbf{c}}_l)\in\widetilde{\mathcal{N}}_i}\exp\left(-\frac{\left\|\Psi_d(\mathbf{c}_i)-\Psi_d(\bar{\mathbf{c}}_l)\right\|^2}{\varepsilon_{\gamma_i}}\right)}, \tag{15}$$

where ε_{γ_i} is defined as the minimal distance between $\boldsymbol{\Psi}_d(\mathbf{c}_i)$ and its nearest neighbor. Consequently, the normalized estimation error is defined as a vector of length d, one for each physical quantity due to its corresponding units:

$$\mathbf{e}(\mathbf{c}_i) = \left[e_i^{(1)}, \ldots, e_i^{(d)}\right], \tag{16}$$

where its jth element is defined by

$$e_i^{(j)} = \frac{|\theta_i^{(j)}(q) - \hat{\theta}_i^{(j)}(q)|}{|\theta_i^{(j)}(q)|}. \tag{17}$$

Note that due to the movement of the sources, the estimation error measures suggested in former works [30,9–11] are incompatible, and thus we suggest an estimation error measure. Our error measure should not be calculated using a norm as the physical quantities composing the parameters vector have different units. Thus, the estimation error is no longer a scalar and each coordinate of the estimation error is calculated individually- one for each of the physical quantities. As a result, the norm signs are redundant and are replaced by absolute values signs. In addition, in order to cancel out the units of each of the individual estimation errors, we normalize the errors using the real values of the parameters, which also adds another virtue- larger absolute differences for a large real parameter value will not be prioritized any longer over smaller absolute differences for a small real parameter value.

We emphasize that the recovery of the unknown location and velocity of a source is determined according to the query point, as pointed out by (14). Based on that point, which we define as the midway point of the trajectory of the source, we minimize the approximation error of the true trajectory by the linear segment, as pointed out by (17).

4.4. Accuracy measure

Following the considerations, mentioned in Section 4.3, of defining the estimation error as a vector of length d, we measure the accuracy of the algorithm as a linear combination of the root mean square error (RMSE) of each one of the elements of (16), given by

$$\text{RMSE} = \sum_{j=1}^{d} \alpha_j \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(e_i^{(j)}\right)^2}, \tag{18}$$

where $\{\alpha_j\}$ are coefficients which represent the significance of the estimation error of each physical quantity, according to user's preference or optimization of the RMSE value (For simplicity, we define as $\{\alpha_j\} = \frac{1}{d}$). Note that in contrast to [9,10,13,11], both sources' movement and even more realistic stationary sources scenarios force us to deal with combination of physical quantities of different units, thus we modify and extend the estimation framework by defining the error as a vector (16), calculating each error component using absolute values (17), and at last a weighted summation of all individual RMSE values in (18).

5. Experimental results

In this section, we demonstrate the capabilities of the proposed passive single-sensor localization algorithm for recovering the location of an acoustic source, in two cases: a baseline case of sources that have a stationary location, and an extensive case of moving sources. In the first stationary subcase, we confirm our choice of an affinity measure between the observations, followed by the recovery of the azimuth angle. In addition, we compare our results to the counterpart results of the prior diffusion maps based single-sensor source localization work [37]. In the second stationary subcase, we extend our experiment for estimation of the radius as well. In the second case, which is the main contribution of this paper, we allow movement of the sources and retrieve their location and velocity. We examine in detail the sensitivity of the proposed algorithm to different hyperparameters (e.g., frame length), variables (e.g., speed) and conditions (e.g., reverberation time) by various experiments.

We describe the simulated setup used for conducting the experimental study, by an efficient implementation [38] of the image method [39]. In all experiments room dimensions were set to $6 \times 5.8 \times 3\,\text{m}^3$, and an omnidirectional microphone was located at $(3, 1, 1.8)$ m. The reverberation time of the room was defined as $T_{60} = 0.3$ s (Except otherwise stated), simulating moderate reverberation conditions. In each location of the source, 1 s (unless else noted) long signal of a zero-mean and unit-variance (for neglecting the system's gain) WGN, sampled at $f_s = 16$ kHz, is transmitted from the source, and after its convolution with the AIR it is measured at the microphone. Consequently, we acquire a total of $m + M$ observations, where m out of them are randomly selected for the training set, while the remaining M samples are allocated for the test set. The corresponding autocorrelation-based feature vector of each observation consists of $D = 800$ lags (if not otherwise specified. The choice of its noted value will be justified later).

5.1. Stationary sources

5.1.1. One-dimensional subcase

Former single-sensor source localization results [10] have been achieved using Mahalanobis distance-based diffusion kernel, focusing on a one-dimensional stationary scenario. On the other hand, various Euclidean distance-based diffusion kernels have been exploited for source localization by multiple-sensors [12,13,18]. Therefore, first of all, and prior to the movement case, we have to test our setting and customized choice of Euclidean distance-based diffusion kernel in a stationary case. We aim to validate the ability of the Euclidean distance-based diffusion kernel to organize the observations according to their azimuth values, in comparison to the results of [10]. In order to do so, a similar stationary experiment has been conducted by positioning all sources at a radius of 1 m from the microphone, whereas their azimuth angles were drawn according to a uniform distribution $U[25^\circ, 64^\circ]$, forming an arc. The experiment was carried out generating training and test sets of 720 observations each.

For considering a metric between the feature vectors which reflects their physical adjacency properly, we compare between two optional distance measures: Euclidean distance, defined by $D_{\text{EUC}}(\mathbf{c}_i, \mathbf{c}_j) = \|\mathbf{c}_i - \mathbf{c}_j\|$; and diffusion distance (13). Fig. 1(a) illustrates a comparison of the aforementioned distance measures according to the azimuth angle, which were measured among all feature vectors with respect to a reference feature vector, representing the minimal azimuth angle (25°). The depicted Euclidean distance is normalized by 1000, for a clear presentation of the behavior of both graphs together. By inspecting the Euclidean distance, we notice it lacks the mandatory property of monotonicty with respect to the azimuth angle, and thus it cannot be used as a distance function between our feature vectors. However, since the Euclidean distance holds monotonicity in the range of approximately 1.85° from the reference angle, we deduce it is adequate for short arcs. Furthermore, in a broader perspective, we can conclude the adequacy of the Euclidean distance is limited to the vicinity of each feature vector, only. Nevertheless, the Euclidean distance can be exploited as part of our diffusion kernel, for two reasons. First, by restricting the computation of the Euclidean distance to small neighborhoods only we can benefit from a reliable affinity measure. Second, the Gaussian kernel is characterized by an inherent locality nature due to its scaling parameter.

In contrast to the Euclidean distance, the diffusion distance is characterized by monotonic behavior throughout the entire azimuth range, thus proving it is a suitable metric for quantifying affinity between our feature vectors. In order to produce the diffusion distance, we use $k = 20$ nearest-neighbors and a scaling parameter of $\varepsilon = 1.45 \cdot \text{median}$ for constructing the diffusion kernel. We assume that only the first eigenvector of the embedding is sufficient (i.e., $d = 1$), since the azimuth is the only varying controlling parameter of the system- this decision will be later validated.

The aforementioned comparison results confirm the nonlinearity of the manifold, which is characterized by a relatively flat surface in the vicinity of each observation, such that this surface looks like a linear Euclidean space. Moreover, these results comply with the counterparts of [13,18], which have been acquired by a similar Euclidean distance-based diffusion kernel for a corresponding stationary one-dimensional scenario. However, the results of [13,18] have been obtained by using dual-sensors, a different feature vector and a dissimilar extension for the eigenvector, as described in detail in Section 4.2.

Fig. 1(b) elaborates on the diffusion mapping of the observations by illustrating the first eigenvector ψ_1 as a function of the azimuth angle. In other words, it shows a comparison of observation representations between two parametric spaces: the one-dimensional location of the sources (i.e., azimuth) and their corresponding independent controlling parameter in the embedded manifold (i.e., ψ_1). We witness that the diffusion mapping follows the azimuth successfully (even linearly for most of the range), up to a monotonic distortion. As a result, the diffusion mapping is capable of accurately discovering the underlying independent parameter dominating the system, that is the location of the source. Moreover, since only the first eigenvector of the embedding is sufficient for monotonic organization of the observations with respect to the azimuth, the choice of $d = 1$ for estimation of the diffusion distance is approved. Hence, the diffusion distance, which approximates the geodesic distance, is accomplished by measuring the distances along the manifold correctly, as well as quantifying the physical adjacency between the locations of the sources. These results support the embeddings results, which originated from similar one-dimensional scenarios (yet following the aforementioned fundamental framework differences) that were presented in [10,12,13,18]- all reflect monotonicity with respect to the azimuth. Note that the latter three, which are based on multiple sensors, do not refer to the extended entries of the eigenvector in their provided embedding results. Our embedding results, on the other hand, prove our suggested extension is accurate, following the arguments mentioned in Section 4.2.

By interpolating the location of the $\tilde{k} = 3$ nearest training neighbors on the manifold, we have established a minimal RMSE of 0.094° in estimating the locations of the test set. For comparison, the single-sensor source localization algorithm presented in [10], which consists of a Mahalanobis distance-based diffusion kernel, has obtained minimal RMSE of nearly 1.2° in recovering the azimuth out of 60 possible predefined (up to perturbations) angles by using 480 training observations (excluding additional observations for the covariance matrix). Both RMSE values were calculated according to the unnormalized version of (17), as proposed in [10]: $e_i = |\theta_i - \hat{\theta}_i|$. We summarize our results in Table 1 and compares them to their counterpart by [10], along with fundamental settings differences.

5.1.2. Two-dimensional subcase

Following the confirmation of the choice of an affinity measure between the observations, the successful azimuth angle recovery, and the comparison to known counterpart results from the literature [10], we can add a second degree of freedom to our system. We aim to challenge the proposed algorithm in recovering the two-dimensional location of stationary sources, where both their radius and azimuth are unknown. In order to do so, we first examine the feasibility of the diffusion framework in organizing the observations monotonically, according to the values of the two hidden independent controlling parameters of the system. We arrange our stationary sources at azimuth angles according to a uniform distribution $U[25, 64]^\circ$. However, in contrast to the former subcase, we allow variability of the radius values according to a uniform distribution $U[1, 1.3]$ m. The experiment was carried out by acquiring training and test sets of 2880 observations each.

We use $k = 10$ nearest-neighbors and a scaling parameter of $\varepsilon = 1.38 \cdot \text{median}$ for constructing the kernel. Since we have added a second degree-of-freedom to the system, we assume that only the first two eigenvectors of the embedding are adequate (i.e., $d = 2$). Fig. 2 depicts the diffusion mapping of the observations of the entire dataset into the embedded manifold. The coloring patterns of the embedded observations, color-coded according to their radius and azimuth values, behave monotonically by each of the location coordinates, thus implying the diffusion mapping perceives the latent parameters dominating the system.

Subsequently, we recover the unknown location of the test observations, by interpolation, using $\tilde{k} = 2$ nearest training neighbors on the manifold. We have established a minimal RMSE of 0.0105, which consists of 0.0041 for radius and 0.017 for azimuth. Furthermore, the localization performance of the proposed algorithm is portrayed in Fig. 3, by a polar plot of 35 sources, randomly picked from the test set.

5.2. Moving sources

In this section, we delve into details and describe by various experiments the different hyperparameters (e.g., frames length), variables (e.g., speed) and conditions (e.g., reverberation time) influencing the performance of the proposed algorithm for the case of moving sources.

For isolating the different impacts of the elements composing the movement of the sources, as well as the factors affecting the performance of the localization algorithm under various circumstances, we simplify the scenarios by focusing on a sole degree-of-freedom system in all of our movement scenarios (i.e., d = 1). Not only it supports conclusions which are relevant regardless of the amount of degrees-of-freedom and even viable for more sophisticated cases, but it also allows us working with significantly smaller dataset compared to scenarios of higher degrees-of-

Table 1
One-Dimensional Stationary Case-Comparison.

Parameter	Ours	Talmon et al. [10]
Diffusion Kernel Type	Euclidean distance-based	Mahalanobis distance-based
Training Set Size	720	480 (5280, including local observations)
Angles Range[°]	39	60
Azimuth Angles Allocation	Uniform distribution over the entire range	60 predefined angles, 1° apart
RMSE[°], by $e_i = \|\theta_i - \hat{\theta}_i\|$	0.094	1.2

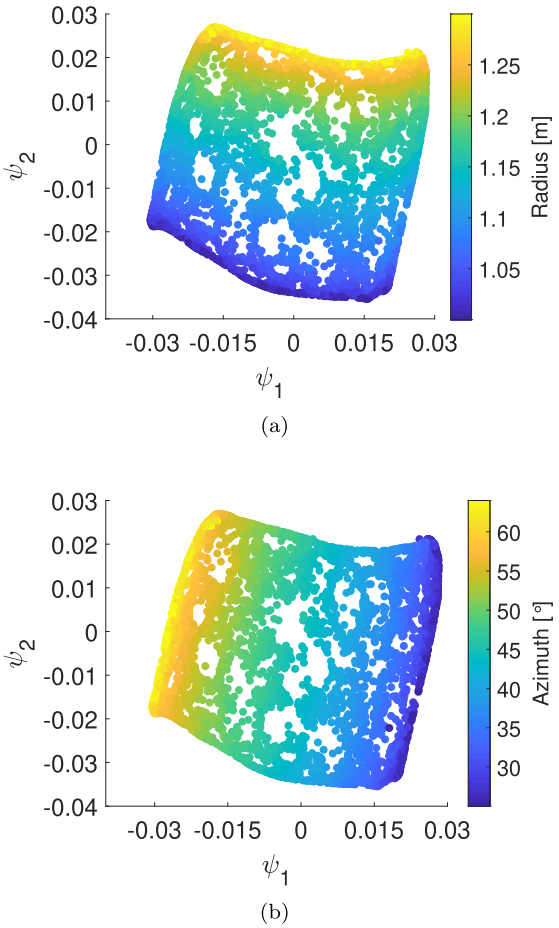

Fig. 2. Diffusion mapping of the observations of the entire dataset into the embedded manifold, color-coded according to (a) radius values, and (b) azimuth values.

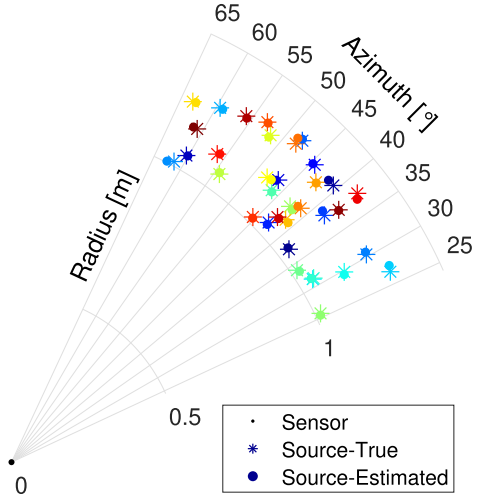

Fig. 3. Polar plot of 35 randomly selected sources out of the test set, in a specific sector. The radial axis represents the radius from the sensor, and the azimuth axis represents the azimuth angles with respect to the sensor. The black dot denotes the location of the sensor, the colored asterisk and disk pairs denote the true and estimated locations, respectively, of each source.

freedom. Scenarios where all parameters of the location and velocity are independent are out of the scope of this work.

Even though only a single controlling parameter is independent, it dictates variations of another two parameters along the movement of the sources, such that it results in an estimation error in each of them. Since the relation between the errors of the independent parameter and the dependent parameters is nonlinear, the errors of the dominated parameters are calculated explicitly, and consequently are taken into consideration at the calculation of the total RMSE.

We assume, for simplicity, that the sources move linearly. In order not to exhaust the reader with dozens values of $k, \tilde{k}$ and ε, due to several simulations composing each movement experiment, we note that both k and $\tilde{k}$ range from 10 to 50 neighbors ($\tilde{k} \leqslant k$), while the proportion of ε to the median ranges from 1.41 to 1.45.

5.2.1. Sensitivity to speed

We examine the accuracy of the proposed localization algorithm with respect to the speed of the moving sources. For that purpose, we position all sources at a distance of 1 m from the microphone and at an azimuth angle of 45°. Their movement is initialized in directions that are drawn according to a uniform distribution $U[45, 85]^\circ$, whereas the speed of all sources is changed in each simulation, in the range of 0.0625 to 1 m/s.

Fig. 4 depicts the total estimation error, as well as the individual estimation errors of the controlling parameter of the scenario (i.e., direction) and the dominated ones (i.e., radius and azimuth). Due to the high variability of the direction its estimation error dictates the behavior of the total error throughout the entire experiment. We yield a high estimation error of the course for slower speed values thanks to the struggle of perceiving variations between the different directions during a bare movement. That struggle, in turn, results in a more scattered embedding (in contrast to a "fine" embedding, as in Fig. 1(b) for example) and thus in a challenge for distinguishing correctly between the various direction values. As the speed gets faster, the accuracy of the estimated direction improves due to a more meaningful movement by the sources, up to a point where the accuracy starts decreasing. That decrease occurs for fast sources and is caused by two factors. First, greater speeds lead to sparser, yet clustered manifolds. That, in turn, makes the estimation of the direction, based on interpolation of the nearest neighbors, more challenging- meaning the variations in the movement of the sources are too fast to be distinguished. In addition, the quasi-stationarity assumption is revoked, hence our autocorrelation-based feature vector, implemented by MATLAB's 'xcorr' function, is no longer valid as the function implicitly assumes WSS signal.

Regrading the radius and the azimuth, as the speed gets faster, a wider range of possible location values is obtained, which consequently results in larger estimation errors.

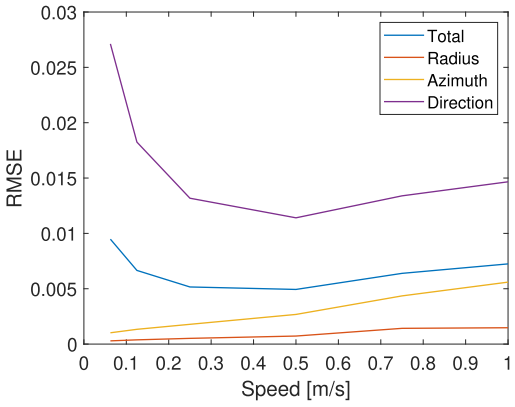

Fig. 4. Performance of the proposed algorithm as a function of the speed of the sources.

5.2.2. Sensitivity to direction

We examine the accuracy of the proposed localization algorithm with respect to the direction of movement of the sources. For that purpose, we position all sources at a distance of 1 m from the microphone and at an azimuth angle of 45°. The movement of the sources is initialized with speed values which are drawn according to a uniform distribution $U[0.25, 0.5]$ m/s, at the examined direction value. The simulation is repeated each time with a different direction of movement of all sources, varying from 5° to 90°, using training and test sets of 720 observations each.

Fig. 5 describes the total estimation error, as well as the individual estimation errors of the independent speed parameter and the dominated radius and azimuth. Not only the speed dictates the behavior of the total estimation error throughout the entire experiment due to its high variability, but its estimation error graph is also followed by the radius estimation error.

As for the azimuth, its estimation error is minimal for a direction of 45° since the azimuth is identical to the direction through the entire movement by the sources, which leads to a degenerated scenario where there is no variability in the azimuth values. As the direction gets away from 45°, the range of possible azimuth values becomes bigger and thus its estimation error aggravates with respect to the gap from the direction of 45°.

The inconsistent behavior of speed's estimation error can be explained by a varying extent of spatial aliasing throughout the experiment, due to the changing direction and the symmetry of the room. Accordingly, it results in an arbitrary allocation of nearest neighbors, despite the true physical adjacency, which translates to a difficulty of distinguishing between sources.

5.2.3. Sensitivity to signal-to-noise ratio

We examine the performance of the proposed localization algorithm with respect to the signal-to-noise ratio (SNR). In order to do so, we repeat the setting of the speed experiment with a slight change, by fixing the speed of all sources to 0.5 m/s. In addition, in each individual simulation, we introduce additive white Gaussian noise (AWGN) of a specific variance value to all signals received by the microphone. The simulation is repeated each time with a varying degree of SNR, ranging from 0 to 30 dB, followed by a scenario free of noise.

Fig. 6 illustrates the total accuracy of the proposed algorithm for various SNR conditions. As expected, as the conditions get harsher, the estimation error grows. Note that although the estimation error is significantly high for SNR of 0 dB, it is not order-of-magnitude higher compared with the other conditions. The reason for that is the use of WGN as a speech signal, thus all frequencies of the transfer function come to realize in the signals received by the microphone, and as a result, there is no a fundamental difference between the embedded manifolds of the various conditions. In addition, from SNR of 20 dB onward the accuracy improves insignificantly and tends to the accuracy of the case of no noise at all (i.e., SNR$\rightarrow \infty$).

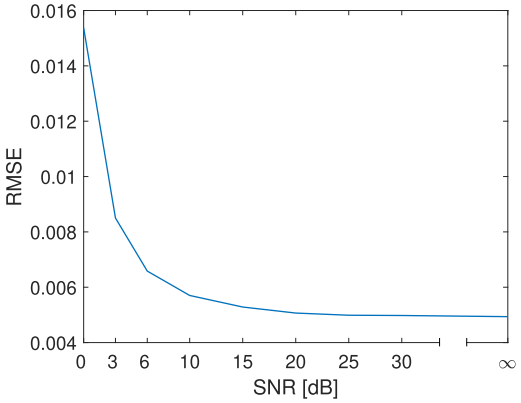

Fig. 6. Performance of the proposed algorithm as a function of SNR.

5.2.4. Sensitivity to reverberation time

We examine the performance of the proposed single-sensor localization algorithm with respect to the reverberation time. For this purpose, we repeat the setting of the speed experiment with a slight change, by fixing the speed of all sources to 0.5 m/s. The simulation is executed each time with a different reverberation time, ranging from 0.128 to 1 s.

Fig. 7 depicts the total estimation error of the algorithm for various reverberation time values, in the blue graph. The estimation errors are displayed in dB for emphasizing the small yet significant differences between close estimation error values. For nearly no reverberation the estimation error is colossal. As the reverberation time gets longer the accuracy improves gradually, up to a point of moderate reverberation (0.4 s), where the performance of the algorithm starts deteriorating steadily. These results validate the implied hypothesis of our work. In contrast to source localization using multiple sensors, where the recovery of the location is based on the direct propagation path of the signal and the reflections are in fact its Achilles' heel, these reflections are essential for source localization using a sole sensor. In addition, by relying solely on the direct direction of arrival, the single-sensor algorithm is incapable of distinguishing between the different locations and velocities of the sources, due to lack of information needed for an accurate estimation. Traditionally, the lack of information is

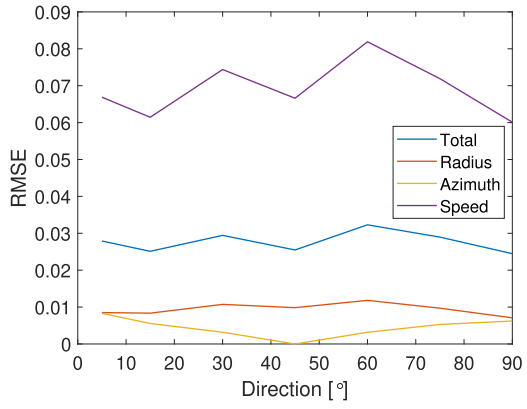

Fig. 5. Performance of the proposed algorithm as a function of the direction of movement of the sources.

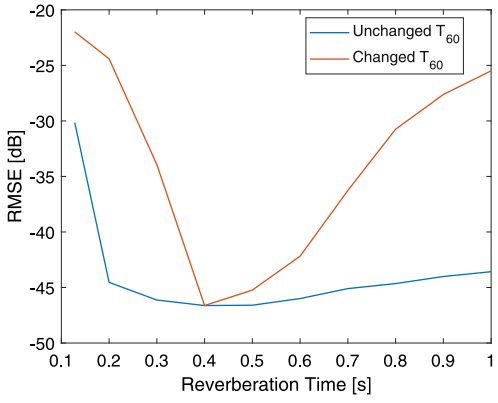

Fig. 7. Performance of the proposed algorithm as a function of reverberation time when both training and test stages share the same T_{60} (blue), and when trained with $T_{60} = 0.4$ s (red).

resolved by adding sensors. However, by exploiting these reflections we can tackle this inadequacy and compensate for the inherent flaw of the single-sensor. On the other hand, similarly to source localization using multiple sensors, the performance of a sole sensor suffers from over-reverberation, but more robustly.

5.2.5. Sensitivity to environmental conditions changes

An outcome of the last experiment provides us the opportunity to inspect the influence of environmental conditions changes, that take place between the training and the test stages, on the performance of the proposed algorithm, as can be seen in the red graph of Fig. 7. In the first stage, we generate a training set with fixed reverberation time of 0.4 s, followed by generation of several test sets, where each one is according to a different reverberation time. First, as the reverberation level of the test stage gets away from reverberation time of 0.4 s, which has prevailed during training, the estimation error worsens. The second finding derives from a comparison between the estimation errors obtained in this experiment and the estimation errors acquired when ideally both training and test stages are held under the same reverberation conditions (as in the previous experiment). We notice that as the absolute difference in the reverberation time between the stages grows, the error deviates from the values achieved ideally- unless the reflections do not play a role in practice (e.g., as at 0.128 s).

From a manifold point of view, small changes in reverberation time results in moderate influence on its structure in general, whereas considerable changes lead to a significant impact on the structure of the manifold. According to the extension, the mapping of the test observations to the embedded manifold is based upon the combination of manifold structure, established by the training observations, and the nearest training feature vectors. Thus, these environmental conditions changes between the training and test stages introduce ambiguities to the extended manifold to an extent. These ambiguities consequently sabotage the efforts for accurate recovery of the unknown locations and velocities of the test set, as achieved when both stages share the same reverberation time. In addition, opposite variations in the reverberation time at the test stage affect significantly different on the estimation results since the single-sensor is more vulnerable to gradual absence of reflections than presence of additional ones. That behavior is reasoned by prevention of essential information for the single-sensor that was previously available during the ideal training stage, due to environmental conditions changes. That absent information is responsible for establishing a manifold structure according to an increasing dominance of the direct propagation path component of the received signals. On the other hand, additional reflections following the environmental conditions changes do not necessarily provide us with more crucial information, compared to what was achieved at the ideal training stage. The reason for that behavior is that reverberation time of 0.4 s is mostly sufficient (in the sense of taking advantage of reflections) for training for single-sensor source localization.

However, note that when the localization is based mostly on direct propagation path of the signal, as at 0.128 s, a conflicting outcome arises where the deviation in the estimation error due to environmental conditions changes is smaller despite the distinct trend. That outcome can be explained by an extensively substantial ambiguity in the original manifold formed ideally, when both training and test stages have shared the same reverberation time. Thus, this significantly ambiguous behavior initially, along with the incremental ambiguity due to environmental conditions changes, leads eventually to a relatively smaller impact on the performance.

5.2.6. Sensitivity to frame length

We examine the performance of the single-sensor localization algorithm with respect to the length of the time frame. All previous experiments refer to a single frame. Thus, the recovery of the location and velocity of the source in these experiments is determined according to a single query point along the source's path, following the discussion and derivation in Section 3.2 and in (18). Rather than executing the algorithm once along the whole path, and consequently carrying an extrapolation throughout the path based on a long linear movement segment (yet may work well for a source that has a constant velocity), we can exploit calculus approach for better estimation results. According to this approach, the whole trajectory of the source, formed by its movement which is characterized by a slow speed and a gradually changing velocity, can be approximated by short linear movement segments. As a result, it allows obtaining a smaller average estimation error by an iterative execution of the algorithm in each frame.

Since the algorithm is repeated and executed in different query points along the path of the source, each point represents a non-overlapping frame (overlapping frames are out of the scope of this work) which is associated with a unique short linear movement segment. Thus, each segment may hold different optimal hyperparameters (i.e., $k, \tilde{k}$ and ε) values for each query point. We define the average RMSE along the path, as an average of the different RMSE values over the frames.

For examining the optimal frame length, we position all sources at a distance of 1 m from the microphone and at azimuth angles, drawn according to a uniform distribution $U[45, 85]^{\circ}$. The movement of the sources is initialized with a speedof 0.5 m/s, at a direction of 45°. The duration of the signal transmitted by each source is set to 2 s. The received signals in each simulation are divided into time frames. The simulation is repeated each time with a different number of non-overlapping frames, varying from one long frame (the whole received signals) to 20 short frames- meaning the frame length varies from 2 s to 0.1 s. The experiment is carried out by acquiring training and test sets of 720 observations each.

Fig. 8 depicts the average total estimation error of the algorithm for various duration values of the segments and their corresponding amount, in dB. On the one hand, the approximation by a linear segment is inaccurate for long frames, unless the source has a constant velocity during the frame. This inaccuracy is expressed by an incremental error, which becomes worse as the frame gets longer- all the more so, when the trajectory of the source is nonlinear. Note that this result is evident even for sources that move linearly, which is the ideal case for such an approximation. Short frames are not good either, regardless of the trajectory of the source, as they are unable to capture the movement properly, and thus incapable of distinguishing between the various azimuth values. We observe the optimal recovery accuracy is achieved by frames of 0.5 s.

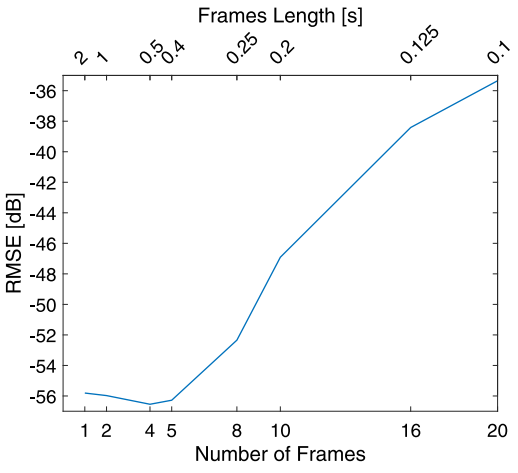

Fig. 8. Performance of the proposed algorithm as a function of the frames length.

6. Conclusions

An unconventional supervised approach for the familiar and mostly ill-posed/underdetermined problem of single-sensor source localization has been presented, using diffusion maps. The proposed algorithm extends manifold learning techniques of former works and demonstrates a proof of concept proposing a state-of-the-art solution for the recovery of the location and velocity of a moving source, using a single microphone. The trajectory, formed by the movement of the source during the frame, is approximated by a linear movement segment. The proposed algorithm implements a data-driven approach for learning the nonlinear structure of the manifold based on the training data. The observations are organized on the manifold according to the location and velocity values of the sources. The unknown location and velocity of a source can be recovered according to its observation's nearest training neighbors on the manifold. The recovery of the location and velocity is determined by the midway point of the segment. Based on that point, we minimize the approximation error of the true trajectory by the linear segment.

Research findings indicate that localization of very slow sources results in good accuracy of the estimated location, at the expense of relatively low accuracy of the estimated direction. Localization of faster sources leads to an improvement of the accuracy of the estimated direction due to a more meaningful movement of the sources, at the expense of the accuracy of the estimated location. The accuracy of the estimated direction starts deteriorating for fast sources, as the variations in their movement are too fast to be distinguished. It is difficult for the algorithm to distinguish between speed values of sources moving in the same direction. The approximation by a linear segment is accurate for frames that are neither too long or short. The results validate the necessity of reflections for yielding an accurate estimation. The algorithm performs well in reverberant and noisy environments, yet is sensitive to environmental conditions changes.

A future study may examine the performance of the algorithm with respect to a violation of its fundamental assumption regarding the characteristics of the velocity. The following topics might be considered as well: a semi-supervised learning scheme, data-driven tracking approaches, and domain adaptation methods to minimize the impact of substantial environmental conditions changes (for example, different enclosures), movement of the sensor, and different domains (including RF signals).

CRediT authorship contribution statement

Eran Zeitouni: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Israel Cohen:** Resources, Writing - review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Eran Zeitouni reports financial support was provided by Israel Science Foundation. Eran Zeitouni reports financial support was provided by National Natural Science Foundation of China. Eran Zeitouni reports financial support was provided by Pazy Foundation. Israel Cohen reports financial support was provided by Israel Science Foundation. Israel Cohen reports financial support was provided by National Natural Science Foundation of China. Israel Cohen reports financial support was provided by Pazy Foundation.

References

[1] Omologo M, Svaizer P. Use of the crosspower-spectrum phase in acoustic event location. IEEE Trans Speech Audio Process 1997;5(3):288–92.
[2] DiBiase JH, Silverman HF, Brandstein MS. Robust localization in reverberant rooms. In: Microphone Arrays. Springer; 2001. p. 157–80.
[3] Schmidt R. Multiple emitter location and signal parameter estimation. IEEE Trans Antennas Propag 1986;34(3):276–80.
[4] Höring H. Comparison of the fixing accuracy of single-station locators and triangulation systems assuming ideal shortwave propagation in the ionosphere. IEE Proceedings F (Radar and Signal Processing) 1990;137 (3):173–6.
[5] Jenkins HH. Small-Aperture Radio Direction-Finding. Artech House on Demand 1991.
[6] Johnson RL, Black Q, Sonsteby AG. HF multipath passive single site radio location. IEEE Trans Aerosp Electron Syst 1994;30(2):462–70.
[7] Poisel R. Electronic Warfare Target Location Methods. Artech House 2012.
[8] Adamy D. EW 103: Tactical Battlefield Communications Electronic Warfare. Artech House 2008.
[9] Talmon R, Kushnir D, Coifman RR, Cohen I, Gannot S. Parametrization of linear systems using diffusion kernels. IEEE Trans Signal Process 2011;60 (3):1159–73.
[10] Talmon R, Cohen I, Gannot S. "Supervised source localization using diffusion kernels," in Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA). IEEE; 2011. p. 245–8.
[11] Laufer B, Talmon R, Gannot S. "Relative transfer function modeling for supervised source localization," in Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA). IEEE; 2013. p. 1–4.
[12] Laufer-Goldshtein B, Talmon R, Gannot S. A study on manifolds of acoustic responses. In: Proc. International Conference on Latent Variable Analysis and Signal Separation. Springer; 2015. p. 203–10.
[13] Laufer-Goldshtein B, Talmon R, Gannot S. Semi-supervised sound source localization based on manifold regularization. IEEE/ACM Trans Audio, Speech, Language Process 2016;24(8):1393–407.
[14] Laufer-Goldshtein B, Talmon R, Gannot S. Manifold-based bayesian inference for semi-supervised source localization. In: Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE; 2016. p. 6335–9.
[15] Laufer-Goldshtein B, Talmon R, Gannot S. Semi-supervised source localization on multiple manifolds with distributed microphones. IEEE/ACM Trans Audio, Speech, Language Process 2017;25(7):1477–91.
[16] Laufer-Goldshtein B, Talmon R, Gannot S. Speaker tracking on multiple-manifolds with distributed microphones. In: Proc. International Conference on Latent Variable Analysis and Signal Separation. Springer; 2017. p. 59–67.
[17] Laufer-Goldshtein B, Talmon R, Gannot S. A hybrid approach for speaker tracking based on TDOA and data-driven models. IEEE/ACM Trans Audio, Speech, Language Process 2018;26(4):725–35.
[18] Laufer-Goldshtein B, Talmon R, Gannot S, et al. "Data-driven multi-microphone speaker localization on manifolds," Foundations and Trends<error l="199" c="Undefined command "/>. Signal Processing 2020;14 (1–2):1–161.
[19] Kupershtein E, Wax M, Cohen I. Single-site emitter localization via multipath fingerprinting. IEEE Trans Signal Processing 2012;61(1):10–21.
[20] Jaffe A, Wax M. Single-site localization via maximum discrimination multipath fingerprinting. IEEE Trans Signal Process 2014;62(7):1718–28.
[21] Khalajmehrabadi A, Gatsis N, Akopian D. Modern wlan fingerprinting indoor positioning methods and deployment challenges. IEEE Commun Surveys Tutorials 2017;19(3):1974–2002.
[22] Wielandt S, Strycker LD. Indoor multipath assisted angle of arrival localization. Sensors 2017;17(11):2522.
[23] Sun X, Gao X, Li GY, Han W. Fingerprint based single-site localization for massive MIMO-OFDM systems. In: Proc. IEEE Global Communications Conference (GLOBECOM). IEEE; 2017. p. 1–7.
[24] Sun X, Gao X, Li GY, Han W. Single-site localization based on a new type of fingerprint for massive MIMO-OFDM systems. IEEE Trans Veh Technol 2018;67(7):6134–45.
[25] Zhang R, Chen G, Zeng Q, Shen L. Single-site positioning method based on high-resolution estimation in VANET localization. IEEE Access 2018;6:54674–82.
[26] Chen L, Yang X, Liu PX, Li C. A novel outlier immune multipath fingerprinting model for indoor single-site localization. IEEE Access 2019;7:21971–80.
[27] Nikoo MS, Behnia F. Single-site source localisation using scattering data. IET Radar Sonar Navig 2017;12(2):250–9.
[28] Bar-Shalom O, Weiss AJ. Transponder-aided single platform geolocation. IEEE Trans Signal Process 2012;61(5):1239–48.
[29] Bar-Shalom O, Weiss AJ. Emitter geolocation using single moving receiver. Signal Processing 2014;105:70–83.
[30] Kushnir D, Haddad A, Coifman R. Anisotropic diffusion on sub-manifolds with application to earth structure classification. Appl Comput Harmonic Analysis 2012;32(2):280–94.
[31] Chung FR, Graham FC. Spectral graph theory. Am Math Soc 1997;92.
[32] Nadler B, Lafon S, Coifman RR, Kevrekidis IG. Diffusion maps, spectral clustering and reaction coordinates of dynamical systems. Appl Comput Harmonic Anal 2006;21(1):113–27.

[33] B. Nadler, S. Lafon, I. Kevrekidis, R.R. Coifman, "Diffusion maps, spectral clustering and eigenfunctions of fokker-planck operators," Advances in neural information processing systems, pp. 955–962, 2006.
[34] Singer A, Coifman RR. Non-linear independent component analysis with diffusion maps. Appl Comput Harmonic Anal 2008;25(2):226–39.
[35] Nyström EJ et al. Über die praktische auflösung von integralgleichungen mit anwendungen auf randwertaufgaben. Acta Math 1930;54:185–204.
[36] Coifman R, Lafon S. Diffusion maps. Appl Comput Harmonic Anal 2006;21 (1):5–30.
[37] Talmon R, Cohen I, Gannot S. "Supervised source localization using diffusion kernels," in Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA). IEEE; 2011. p. 245–8.
[38] E.A. Habets, "Room impulse response generator," Technische Universiteit Eindhoven, Tech. Rep, vol. 2, no. 2.4, p. 1, 2006.
[39] Allen JB, Berkley DA. Image method for efficiently simulating small-room acoustics. J Acoust Soc Am 1979;65(4):943–50.