KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Maja Taseska, Toon van Waterschoot, (2019) **On spectral embeddings for supervised binaural source localization**
Archived version	Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher
Published version	
Journal homepage	http://eusipco2019.org
Author contact	your email maja.taseska@esat.kuleuven.be
IR	

(article begins on next page)

On spectral embeddings for supervised binaural source localization

Maja Taseska and Toon van Waterschoot
{maja.taseska, toon.vanwaterschoot}@esat.kuleuven.be
KU Leuven, Dept. of Electrical Engineering (ESAT-STADIUS/ETC), Leuven, Belgium

***Abstract*—Advances in data-driven signal processing have resulted in impressively accurate signal and parameter estimation algorithms in many applications. A common element in such algorithms is the replacement of hand-crafted features extracted from the signals, by data-driven representations. In this paper, we discuss low-dimensional representations obtained using spectral methods and their application to binaural sound localization. Our work builds upon recent studies on the low-dimensionality of the binaural cues manifold, which postulate that for a given acoustic environment and microphone setup, the source locations are the primary factors of variability in the measured signals. We provide a study of selected linear and non-linear spectral dimensionality reduction methods and their ability to accurately preserve neighborhoods, as defined by the source locations. The low-dimensional representations are then evaluated in a nearest-neighbor regression framework for localization using a dataset of dummy head recordings.**

***Index Terms*—binaural source localization, dimensionality reduction, manifold learning**

I. Introduction

Binaural sound localization consists of estimating the location of a source, using signals captured by microphones at the ear canal entrance of the human auditory system. Although algorithms for hearing aids and humanoid robots are the leading applications, the concepts are equally relevant for localization with arbitrary two-microphone configurations as well [1], [2]. Typically binaural localization starts by extracting spatial cues, such as interaural level, phase, and time differences [2]–[5], with the objective to map these cues to the source Direction-of-Arrival (DOA) that consist of azimuth and/or elevation. Researchers have argued that a data-driven approach is crucial to accurately model the complex relationship between source locations and interaural cues in reverberant environments. Several algorithms based on parametric statistical models, typically Gaussian mixtures, were developed in this line of research [2], [3], [5].

In recent literature, the intrinsic geometric structure of binaural signals was exploited to design different data-driven localization algorithms [1], [6], [7]. The common paradigm is that for a given environment and microphone setup, the source DOA is the primary factor of variability in the binaural signals. As a result, high-dimensional feature vectors constructed from binaural cues are bound to lie near a low-dimensional manifold embedded in the feature space. An ensuing implication is that localization algorithms can be more effective in a suitable low-dimensional domain consistent with the manifold geometry. Note that *dimensionality reduction*, *manifold learning*, and *representation learning* are synonymous in this context. Acoustic source localization approaches that benefit from such geometric insights range from parametric and probabilistic [4], [6], to non-parametric and deterministic [1].

In this paper, we conduct a study of selected spectral methods for dimensionality reduction and their applicability to binaural source localization. The resulting low-dimensional representations, referred to as *spectral embeddings*, are obtained from the eigenvectors of certain symmetric matrices derived from the data [8]. Typical spectral methods include Principal Component Analysis (PCA), Laplacian Eigenmaps (LEM) [9], and diffusion maps [10]. The latter were studied in [1], [7] for source localization with simulated recordings, providing some encouraging results. In this work, we discuss several linear and non-linear spectral methods, emphasizing the importance of non-linearities for accurate localization. Our discussion is supported by experiments with the CAMIL dataset of real dummy head recordings [4], [6]. The paper is organized as follows: in Section II we formalize the non-parametric, supervised binaural source localization problem. In Section III, we discuss the selected spectral embeddings, and in Section IV we present our experimental results.

II. Supervised binaural source localization

Let $s_{\mathrm{l}}(\tau)$ and $s_{\mathrm{r}}(\tau)$ denote short-duration signals captured at the left and right microphones in a reverberant environment, during activity of an acoustic source with arbitrary frequency content. The goal is to estimate the position tuple $r = (\theta, \phi)$ of azimuth and elevation with respect to the microphones. For this study, we assume that a single source is active at a time, and impose no assumptions on the level of background noise.

A. Binaural data model

The first step in the source localization pipeline is transforming time-domain signals $s_l(\tau)$ and $s_r(\tau)$ to a suitable feature vector that preserves the relevant spatial cues. This is generally achieved using time-frequency (TF) transforms

This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of a Postdoctoral Research Fellowship of the Research Foundation Flanders - FWO-Vlaanderen (no. 12X6719N) and KU Leuven Internal Funds C2-16-00449 and VES/19/004. The research leading to these results has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program / ERC Consolidator Grant: SONORA (no. 773268). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.

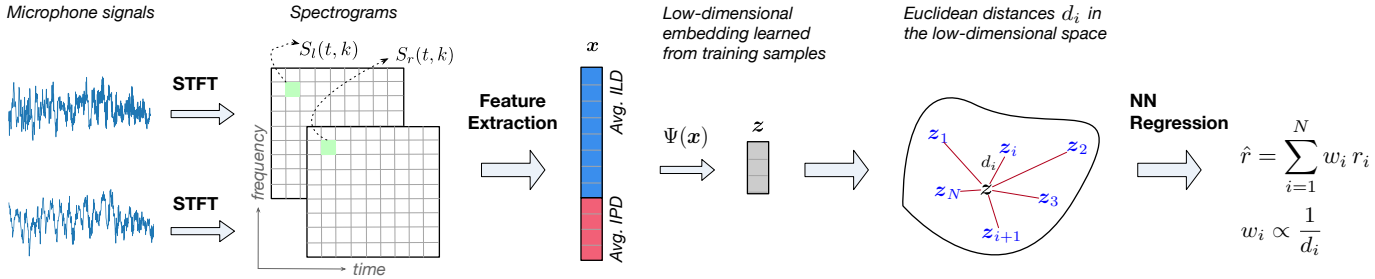


Fig. 1: The supervised source localization pipeline: binaural feature extraction, low-dimensional embedding, and nearest-neighbor regression.

with an appropriate resolution [1], [3], [5], [6]. We consider the Short-time Fourier Transform (STFT) with a length of 512 samples and 50% overlap, at 8 kHz sampling rate. Let $S_l(t,k)$ and $S_r(t,k)$ denote the STFT coefficients of the microphone signals, for a given TF bin (t,k). The relevant spatial cues, namely, the Interaural Level Difference (ILD) α_{tk} and Interaural Phase Difference (IPD) ϕ_{tk} are computed as

$$\alpha_{tk} = 20\log_{10}\frac{|S_l(t,k)|}{|S_r(t,k)|}, \quad \phi_{tk} = \angle\frac{S_l(t,k)}{S_r(t,k)}. \tag{1}$$

Assuming that a single source is active, we compute the time-averaged ILDs and IPDs across T frames, as done in [4]

$$a_k = \frac{1}{T}\sum_{t=1}^{T}\alpha_{tk}, \quad p_k = \frac{1}{T}\sum_{t=1}^{T}\exp(j\phi_{tk}). \tag{2}$$

A feature vector $\boldsymbol{x} \in \mathbb{R}^D$ is obtained by concatenating the ILDs, and the real and imaginary parts of the IPDs in selected frequency ranges $[k_1, k_2]$ and $[k_3, k_4]$, as follows

$$\boldsymbol{x} = [a_{k_1}, \dots a_{k_2}, \mathcal{R}\{p_{k_3}\}, \mathcal{I}\{p_{k_3}\}, \dots, \mathcal{R}\{p_{k_4}\}, \mathcal{I}\{p_{k_4}\}]^{\mathrm{T}}. \tag{3}$$

While IPDs carry reliable location cues at frequencies below 2 kHz [11], ILDs have been shown to contribute to localization accuracy at both low and high frequencies [4]. Therefore, we choose the frequency ranges accordingly: 100 - 2000 Hz for IPDs, and 100 - 3800 Hz for ILDs, resulting in a 481-dimensional feature vector $\boldsymbol{x}$.

B. Non-parametric supervised localization

The framework in our study consists of two stages: an *unsupervised training* stage for learning a low-dimensional representation of $\boldsymbol{x}$, and a *supervised regression* stage for localization. Given a labeled training set of N samples, $\mathcal{T} = \{(\boldsymbol{x}_i, r_i)\}_{i=1}^N$, the two stages are defined as follows:

i) **Training:** Use the samples $\{\boldsymbol{x}_i\}_{i=1}^N$ to learn an embedding map $\Psi : \mathbb{R}^D \to \mathbb{R}^d$, with $d << D$, such that to each $\boldsymbol{x}_i \in \mathbb{R}^D$, a new representation $\boldsymbol{z}_i = \Psi(\boldsymbol{x}_i)$, $\boldsymbol{z}_i \in \mathbb{R}^d$ is assigned. The criteria for a suitable map Ψ shall become clear shortly.

ii) **Localization:** Given a new observation $\boldsymbol{x}$, find its representation $\boldsymbol{z} = \Psi(\boldsymbol{x})$ using the map learned in the training stage. The sound source is then localized by a Nearest-Neighbor (NN) regression in $\mathbb{R}^d$ using the training set labels $\{r_i\}_{i=1}^N$ as follows

$$\hat{r} = \sum_{i=1}^{N} w_i\, r_i, \quad \text{with} \quad w_i = \frac{\|\boldsymbol{z} - \boldsymbol{z}_i\|^{-1}}{\sum_{j=1}^{N}\|\boldsymbol{z} - \boldsymbol{z}_j\|^{-1}}. \tag{4}$$

It is clear from (4) that the map Ψ should preserve neighborhoods as defined by the source locations, while being insensitive to variations caused by noise, reverberation, and spectral content. We note that the choice of regression weights in (4) is not critical: although our weights are inverse-proportional to the distance, exponentially decaying weights as in [1], would not change the conclusions of our study. The complete signal processing pipeline of this section is illustrated in Figure 1.

III. Spectral embeddings under study

Given the standard supervised framework in Section II, our objective is to investigate the ability of different spectral methods to provide a low-dimensional embedding map Ψ with the aforementioned properties. Spectral methods provide Ψ using the eigenvectors of suitable symmetric matrices constructed from the data [8]. We consider PCA and Locality-Preserving Projection (LPP) as representatives from the linear methods, and Laplacian Eigenmaps (LEM), as a non-linear method.

A. Principal components analysis

Classical PCA learns a linear map Ψ using the eigenvectors of the sample covariance matrix of the data [12]. Assuming that the feature vectors in the training set are zero-mean, the covariance matrix is given by

$$\boldsymbol{C} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}}, \quad \boldsymbol{C} \in \mathbb{R}^{D\times D}. \tag{5}$$

Let $\{\boldsymbol{v}_i\}_{i=1}^d$ denote the d largest eigenvectors of $\boldsymbol{C}$. The d-dimensional PCA embedding is given by

$$\boldsymbol{z}_i = \Psi_{\mathrm{pca}}(\boldsymbol{x}_i) = \boldsymbol{A}_{\mathrm{pca}}^{\mathrm{T}}\,\boldsymbol{x}_i, \quad \boldsymbol{A}_{\mathrm{pca}} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_d]. \tag{6}$$

Given a new feature vector $\boldsymbol{x}$ outside of the training set, its embedding $\boldsymbol{z}$ is easily obtained by $\boldsymbol{z} = \boldsymbol{A}_{\mathrm{pca}}^{\mathrm{T}}\,\boldsymbol{x}$.

As the relationship between source locations and binaural feature vectors is typically non-linear, PCA is not well-suited for learning efficient low-dimensional representations of the latter. To better illustrate the relationship of PCA to the spectral methods discussed in the following sections, we note that

PCA leads to the same embedding as metric multidimensional scaling. As such, it provides an embedding that optimally preserves Euclidean distances within the training set [8], [13].

B. Laplacian eigenmaps and diffusion maps

LEM, proposed in [14], belongs to the class of non-linear spectral approaches. In contrast to PCA, where the objective is to optimally preserve Euclidean distances, LEM seeks to preserve similarities as measured by a positive semi-definite, exponentially-decaying kernel function $k : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}$. A theoretically motivated choice for a kernel function k [14] is the isotropic Gaussian kernel, given by

$$k_\varepsilon(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{\varepsilon}\right), \tag{7}$$

where ε is the kernel bandwidth. The choice of ε is data-dependent and can be crucial in determining the neighborhood-preserving properties of the embedding.

The kernel gives rise to an $N \times N$ matrix $\boldsymbol{K}$ with entries $\boldsymbol{K}[i,j] = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$. Let $\boldsymbol{D}$ denote a diagonal matrix with entries $\boldsymbol{D}[i,i] = \sum_{j=1}^{N} \boldsymbol{K}[i,j]$, and let $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{K}$ (this is the graph Laplacian, well-known in spectral graph theory [15]). If $\boldsymbol{Z} \in \mathbb{R}^{N \times d}$ denotes the matrix with rows$\{\boldsymbol{z}_i^{\mathrm{T}}\}_{i=1}^{N}$, LEM embeddings are obtained by minimizing the following cost [14]

$$\mathcal{J}(\boldsymbol{Z}) = \sum_{i,j} \|\boldsymbol{z}_i - \boldsymbol{z}_j\|^2 \, \boldsymbol{K}[i,j] = \mathrm{trace}\{\boldsymbol{Z}^T \, \boldsymbol{L} \, \boldsymbol{Z}\}. \tag{8}$$

The locality-preserving property is clear from (8): points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ with large similarity $\boldsymbol{K}[i,j]$ must be mapped to nearby points $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$. To avoid arbitrary scaling, $\boldsymbol{Z}$ is obtained by including a constraint in the optimization problem as follows

$$\arg\min_{\boldsymbol{Z}} \ \mathrm{trace}\{\boldsymbol{Z}^T \, \boldsymbol{L} \, \boldsymbol{Z}\}, \quad \text{subject to } \boldsymbol{Z}^{\mathrm{T}} \boldsymbol{D} \boldsymbol{Z} = \boldsymbol{I}. \tag{9}$$

The columns of the optimal $\boldsymbol{Z}$ are given by the d smallest eigenvectors of the generalized eigenvalue problem

$$\boldsymbol{L}\boldsymbol{\psi} = \lambda \, \boldsymbol{D} \, \boldsymbol{\psi}. \tag{10}$$

If $\{\boldsymbol{\psi}_i\}_{i=1}^{N}$ denote the eigenvectors, with eigenvalues $0 = \lambda_1 < \lambda_2 \leq \ldots, \leq \lambda_N$, the d-dimensional LEM is given by the smallest eigenvectors (excluding $\boldsymbol{\psi}_1$ which is constant) [14]

$$\boldsymbol{z}_i = \Psi(\boldsymbol{x}_i) = \left[\, \boldsymbol{\psi}_2[i], \ \boldsymbol{\psi}_3[i], \ \ldots, \ \boldsymbol{\psi}_{d+1}[i] \right]^{\mathrm{T}}. \tag{11}$$

It is straightforward from (10) that the embedding can instead be obtained from the largest eigenvectors of $\boldsymbol{D}^{-1}\boldsymbol{K}$. In fact, this approach is preferred when implementing LEM with iterative eigenvector solvers [16], and is also used in the related diffusion maps algorithm [10]. Therefore we solve for the eigenvectors of $\boldsymbol{D}^{-1}\boldsymbol{K}$ in our experiments as well.

A crucial difference between PCA, and non-linear spectral methods lies in the extension of the embedding to points outside of the training set. To extend LEM without solving a new eigenvalue problem, the embedding $\boldsymbol{z}$ of a new feature vector $\boldsymbol{x}$ is obtained as a linear combination of the embeddings $\{\boldsymbol{z}_i\}_{i=1}^{N}$ from the training set [17]

$$\boldsymbol{z} = \Psi(\boldsymbol{x}) = \sum_{i=1}^{N} \frac{k_\epsilon(\boldsymbol{x}, \boldsymbol{x}_i)}{\sum_{j=1}^{N} k_\epsilon(\boldsymbol{x}, \boldsymbol{x}_j)} \, \boldsymbol{z}_i. \tag{12}$$

C. Locality-preserving projection

Linear embeddings can also be constructed with a neighborhood-preserving property. This notion is used in Locality-Preserving Projection (LPP), by minimizing a cost function similar to (8), while constraining $\boldsymbol{z}_i$ to be a linear transformation of $\boldsymbol{x}_i$ [18]. The LPP map is obtained using the eigenvectors of the following problem [18] (notice the similarity to the LEM problem in (10))

$$\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\psi} = \lambda \boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\psi}. \tag{13}$$

The smallest eigenvectors $\{\boldsymbol{\psi}_i\}_{i=1}^{d}$ are the columns of LPP map $\boldsymbol{A}_{\mathrm{lpp}}$ and the d-dimensional embedding is given by

$$\boldsymbol{z}_i = \Psi_{\mathrm{lpp}}(\boldsymbol{x}_i) = \boldsymbol{A}_{\mathrm{lpp}}^{\mathrm{T}} \, \boldsymbol{x}_i, \quad \boldsymbol{A}_{\mathrm{lpp}} = [\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \ldots, \boldsymbol{\psi}_d]. \tag{14}$$

LPP is attractive due to its linearity, allowing for easy extension to new samples, and its neighborhood-preserving property inherited by the LEM-like cost function.

IV. Experimental analysis

To evaluate the spectral embeddings, we used the CAMIL dataset [4] of binaural recordings made with a dummy head, mounted onto a turntable system in a reverberant room. The source location is fixed at 2.7 m distance from the dummy head, and signals are recorded for 10.800 pan-tilt states of the dummy head, corresponding to source azimuth in the range $[-180°, 180°]$, and elevation in the range $[-60°, 60°]$, with $2°$ resolution. To provide meaningful results, realistic mismatch between training and test sets must be ensured, including differences due to source spectral content, random noise, and source location. Therefore, we used white noise recordings (1 s per recording) for training and speech utterances from the TIMIT database (1-5 s per recording) for testing. In addition, white noise was added to the speech signals (Signal-to-Noise Ratio (SNR) specified in the experiments below). Training was done with only 50% of the pan-tilt states, which amounts to sampling of the azimuth-elevation space with $4°$ resolution. In this manner, we ensure that the test set contains DOAs that are not encountered during training.

In all experiments, we used a kernel bandwidth of $\varepsilon = 0.1$. Note that the choice of bandwidth is an interesting research question by itself, and is often influenced by the sampling distribution of data in the training set.

A. Experiment 1: Neighborhood analysis

In this experiment, we evaluate the ability of PCA, LPP, and LEM to preserve neighborhoods in terms of the source location, as a function of the number of neighbors. White noise with SNR of 10 dB was added to the test signals. Given a low-dimensional embedding Ψ, the experiment consists of

i) For each test point $\boldsymbol{x}$ and its embedding $\boldsymbol{z} = \Psi(\boldsymbol{x})$, compute the Euclidean distances $\|\boldsymbol{z} - \boldsymbol{z}_i\|$, from training points $\boldsymbol{z}_i \in \mathcal{T}$. Find the M nearest neighbors as defined by these distances.

ii) If θ and ϕ are the azimuth and elevation of the test point, consider the azimuths $\theta_1, \ldots, \theta_M$ and elevations $\phi_1, \ldots, \phi_M$ of the neighbors. Compute the mean absolute differences

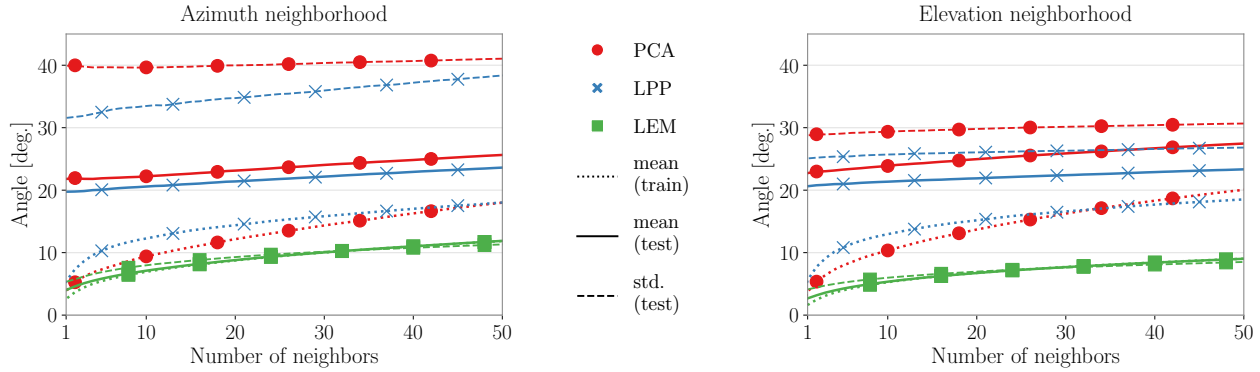


Fig. 2: Results from Experiment 1: average distance within neighborhoods for the different spectral embeddings.

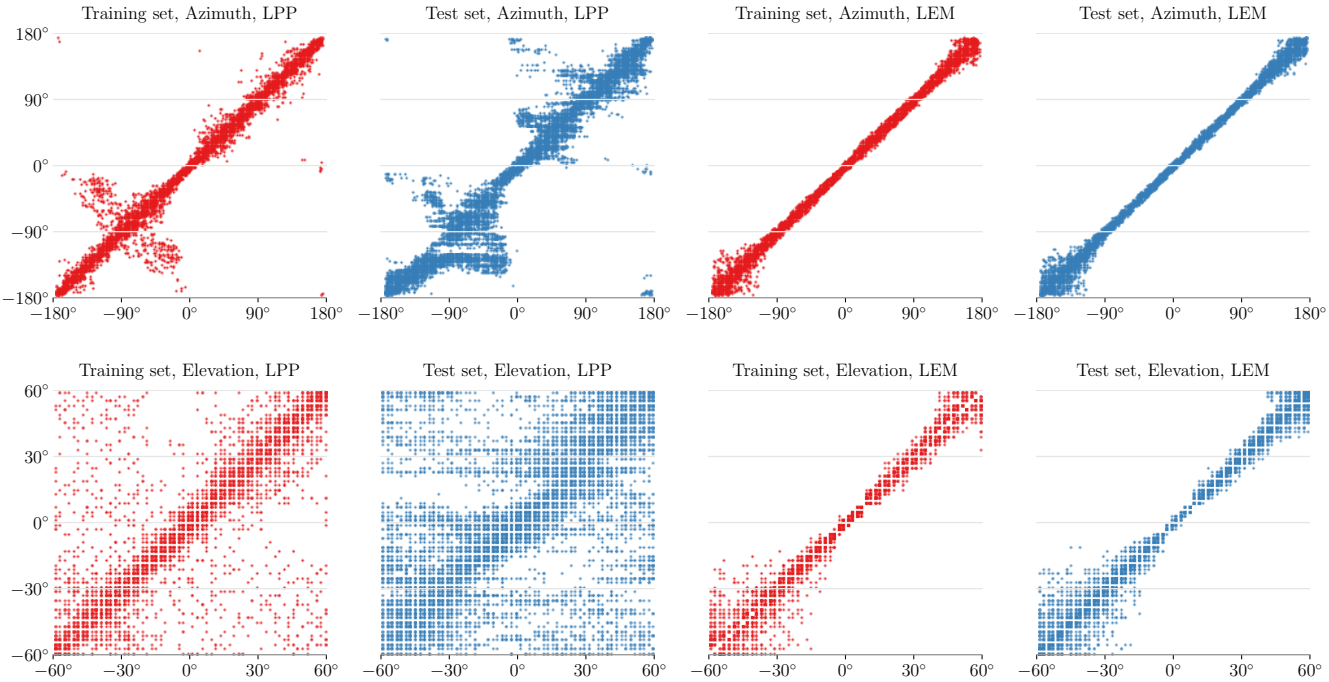


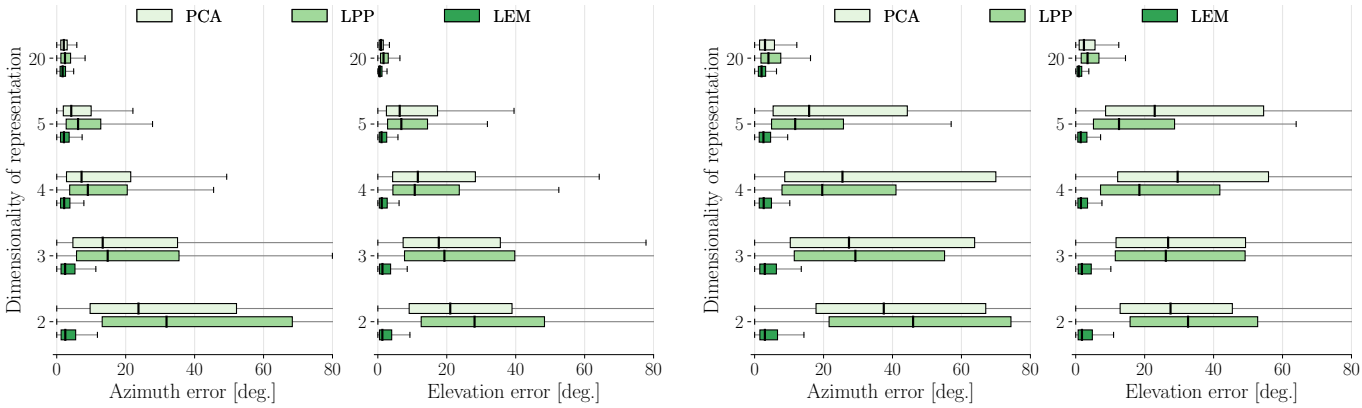
Fig. 3: True locations of sources in the training/test set plotted versus the locations of their respective nearest neighbor in the training set.

$\Delta_{\theta,\boldsymbol{x}} = \sum_{i=1}^{M} M^{-1}|\theta - \theta_i|$ and $\Delta_{\phi,\boldsymbol{x}} = \sum_{i=1}^{M} M^{-1}|\phi - \phi_i|$. iii) Finally, compute the mean and the standard deviation of $\Delta_{\theta,\boldsymbol{x}}$ and $\Delta_{\phi,\boldsymbol{x}}$ across the 10800 samples $\boldsymbol{x}$ from the test set.

The procedure is applied to Ψ_{pca}, Ψ_{lpp}, and Ψ_{lem}, for $M \in [1, 50]$, and the results for 4-dimensional maps are shown in Figure 2, (we observed similar trends for all tested $d \in [2, 20]$). The superior neighborhood preservation of LEM over the linear methods is demonstrated for both azimuth and elevation. Notably, LEM exhibits similar behavior on the test set as on the training set. Moreover, the standard deviation similar to the sample mean indicates absence of outliers and consistent behavior independent of the source location. None of these desirable properties are observed for the linear methods. We note however that LPP provides smaller neighborhood variance than PCA, as expected from its cost function. A qualitative illustration of the neighborhoods is provided by the scatter plots in Figure 3. Each point corresponds to one sample from the training or test set, with the true source angle in the x-axis, and the angle of its first nearest neighbor from the training set in the y-axis. These plots provide the useful insight that performance deteriorates for broadside azimuth and large (positive or negative) elevation angles.

B. Experiment 2: Localization accuracy

In this experiment, PCA, LPP, and LEM with dimensionality of 2, 3, 4, 5, and 20, were employed for supervised source localization using the framework described in Section II. Each of the 10800 speech samples in the test set was localized for SNR of 10 dB and 0 dB. To capture the error variability across the dataset, we illustrate the results using box-plots in Figure 4. Although PCA and LPP approach the performance of LEM as the dimensionality increases beyond 20, the advantage of LEM over the linear methods is remarkable. In particular, 2-dimensional LEM result in a median azimuth error of 4.6°



(a) Signal-to-noise ratio of 10 dB

(b) Signal-to-noise ratio of 0 dB

Fig. 4: Localization accuracy of the spectral embeddings for different dimensionality. The figure illustrates median error (thick line), the first and third quartiles (boxes), and the interquartile ranges (whiskers).

and 4.2°, and median elevation error of 3.1° and 1.8°, for the different SNRs. The same error figures for PCA and LPP exceed 50° for azimuth, and 40° for elevation. The ability to recover source locations with 2-dimensional spectral embeddings has been suggested in previous research [1], [6]. Our experiments support this finding for LEM as well, while indicating that localization accuracy can be further improved by increasing the dimensionality. In particular, the accuracy for near end-fire azimuth, and large elevation could benefit from a larger dimensionality. The results demonstrate a clear relationship between the neighborhood-preserving properties of an embedding and its robustness to noise. Besides the remarkable noise robustness of LEM, we note that while PCA outperforms LPP at moderate SNR, the neighborhood-preserving properties of LPP make the latter more accurate than PCA at low SNRs.

V. Conclusions

In this paper, we studied spectral embeddings in the context of supervised binaural source localization. We showed that although linear methods can be applied for embedding with sufficiently high dimensionality, they are unable to capture the mapping from binaural cues to source locations with only a few coordinates. Non-linearity and neighborhood-preservation emerged as two key prerequisites if low-dimensionality, accuracy, and robustness to noise are desired. Starting from these insights, the applicability of spectral embeddings needs to be investigated in time-varying acoustic conditions. In particular, the challenging problem is to find a robust embedding which remains consistent with the source DOA, even when the acoustic channels in the test set are significantly different from those in the training set.

References

[1] B. Laufer, R. Talmon, and S. Gannot, "Relative transfer function modeling for supervised source localization," in *IEEE Work. Appl. Signal Process. to Audio Acoust. (WASPAA)*, 2013, pp. 1–4.

[2] M. Mandel, R. Weiss, and D. Ellis, "Model-based expectation-maximization source separation and localization," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 2, pp. 382–394, 2010.

[3] T. May, S. Van De Par, and A. Kohlrausch, "A probabilistic model for robust localization based on a binaural auditory front-end," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 19, no. 1, pp. 1–13, 2011.

[4] A. Deleforge, F. Forbes, and R. Horaud, "Acoustic space learning for sound source separation and localization on binaural manifolds," *Int. J. Neural Syst.*, vol. 25, no. 1, 2015.

[5] J. Woodruff and D. L. Wang, "Binaural localization of multiple sources in reverberant and noisy environments," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 20, no. 5, pp. 1503–1512, 2012.

[6] A. Deleforge and R. Horaud, "2D sound-source localization on the binaural manifold," in *IEEE Int. Work. Mach. Learn. Signal Process. MLSP*, 2012.

[7] B. Laufer-Goldshtein, R. Talmon, and S. Gannot, "A Study on Manifolds of Acoustic Responses," in *Int. Conf. Latent Var. Anal. Signal Sep. (LVA/ICA)*, 2015, pp. 203–210.

[8] S. Lawrence K., W. Kilian Q., S. Fei, H. Jihun, and L. Daniel D., "Spectral Methods for Dimensionality Reduction," in *Semi-Supervised Learn.*, O. Chapelle, B. Schölkopf, and A. Zien, Eds. MIT Press, 2006, ch. 16, pp. 293–308.

[9] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 6, no. 15, pp. 1373–1396, 2003.

[10] B. Nadler, S. Lafon, R. Coifman, and I. G. Kevrekidis, "Diffusion maps - A probabilistic interpretation for spectral embedding and clustering algorithms," in *Lect. Notes Comput. Sci. Eng.*, A. N. Gorban, B. Kegl, D. C. Wunsch, and Z. Andrei, Eds. Springer-Verlag Berlin Heidelberg, 2008, ch. 10, pp. 238–260.

[11] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT Press, 1997.

[12] I. T. Jolliffe, *Principal Component Analysis. Second Edition*, 2002.

[13] T. Cox and M. Cox, *Multidimensional Scaling*. Chapman & Hall, 1994.

[14] M. Belkin, "Problems of Learning on Manifolds," Ph.D. dissertation, The University of Chicago, 2003.

[15] F. R. K. Chung, *Spectral Graph Theory*. Providence, RI: American Mathematical Society, 1997.

[16] R. B. Lehoucq, D. C. Sorensen, and C. Yang, *ARPACK users' guide: solution of large-scale eigenvalue problems with implicitly restarted Arnoldi methods*. SIAM, Software, Environments, and Tools, 1998.

[17] Y. Bengio, J.-F. Paiement, P. Vincent, O. Delalleau, N. Le Roux, and M. Ouimet, "Out-of-sample extensions for LLE, Isomap, MDS, Eigenmaps and Spectral Clustering," *Adv. Neural Inf. Process. Syst.*, vol. 16, pp. 177–184, 2004.

[18] X. He and P. Niyogi, "Locality preserving projections," in *Proc. 16th Int. Conf. Neural Inf. Process. Syst.*, 2003, pp. 153–160.