EUSIPCO 2013 1569744627

RELATIVE TRANSFER FUNCTION IDENTIFICATION ON MANIFOLDS FOR SUPERVISED GSC BEAMFORMERS

Ronen Talmon

Department of Mathematics
Yale University, U.S.A.

Sharon Gannot

Faculty of Engineering
Bar-Ilan University, Israel

ABSTRACT

Identification of a relative transfer function (RTF) between two microphones is an important component of multichannel hands-free communication systems in reverberant and noisy environments. In this paper, we present an RTF identification method on manifolds for supervised generalized sidelobe canceler beamformers. We propose to learn the manifold of typical RTFs in a specific room using a novel extendable kernel method, which relies on common manifold learning approaches. Then, we exploit the extendable learned model and propose a supervised identification method that relies on both the a priori learned geometric structure and the measured signals. Experimental results show significant improvements over a competing method that relies merely on the measurements, especially in noisy conditions.

Index Terms— Array signal processing, system identification, manifold learning, acoustic modeling

1. INTRODUCTION

Identification of a relative transfer function (RTF) between two microphones is an important component of multichannel hands-free communication systems in reverberant and noisy environments. Modern beamformers often require an estimate of the room impulse response (RIR) relating the source and the microphones. Room impulse responses (or their respective acoustic transfer function (ATF)) estimation in a noisy environment is a cumbersome task. Hence, it was proposed [1] to replace the ATFs by the RTFs in the beamformer design. An accurate identification of the RTFs leads to significant improvement in the performance of the beamformer [1, 2]. Over the years, several RTF identification methods based merely on the measured signals have been proposed [3, 4]. However, these methods often suffer from poor results in noisy environments, since a large number of independent parameters need to be estimated. To overcome such a shortcoming in adverse environments, recent supervised *system* identification methods [5, 6] utilize representative RIRs to form a model in advance, which, in turn, is incorporated into the identification procedure.

In this paper, we suggest a different, supervised, approach to the problem of RTF identification, similar to [7, 8]. Although acoustic modeling is considered a difficult problem, we observe that RIRs are governed by few parameters, e.g. the size and geometry of the room, the positions of the source and the microphones, and the reflective properties of the walls. As a consequence, acoustic paths exhibit geometric structures of low dimensionality, which are often called *manifolds*, and may be accurately parameterized using manifold learning methods [5, 6, 9]. In this work, we consider a room with a pair of microphones in a fixed location and assume that the possible positions of the desired source are confined to a specific known region. Examples to such rooms are a conference room, in which the microphone array is located in a fixed location on the conference room table and the speakers sit around the table in designated locations, or an office, in which the microphone array in located in a fixed location on the desk or on the computer screen and the speaker sits behind the desk in typical positions. Focusing on designing a solution for a specific room enables us to assume the availability of a training set, i.e., a set of RTFs from the region of possible source positions. Such a set may be acquired by performing repeated recordings of fully-exciting training signal from the region of source positions in controlled noiseless conditions, and the corresponding RTFs can be accurately estimated using a standard system identification method, e.g. a least squares fit. We learn the manifold of the training set of the RTFs using a novel extendable kernel method [10], which relies on Laplacian Eigenmaps [11] and Diffusion Maps [12]. Then, we utilize the extendable learned model and propose a supervised identification method that relies on both the a priori learned geometric structure and the measured signals. Experimental results show significant improvements over a competing method that relies merely on the measurements, especially in noisy conditions.

2. RTF IDENTIFICATION

In this section, we repeat the procedure presented in [1]. Let $s(n)$ denote a speech signal, and let $u(n)$ and $w(n)$ denote stationary measurement noise signals. The signals are mea-

sured by two microphones:

$$x(n) = h_1(n) * s(n) + u(n)$$
$$y(n) = h_2(n) * s(n) + w(n) \quad (1)$$

where $*$ represents convolution, and $h_1(n)$ and $h_2(n)$ are the RIRs between the two microphones and the source, respectively. An equivalent representation of (1) is

$$y(n) = h(n) * x(n) + v(n)$$
$$v(n) = w(n) - h(n) * u(n) \quad (2)$$

where $h(n)$ represents the *relative* impulse response between the microphones with respect to the source and satisfies $h_2(n) = h(n) * h_1(n)$. In (2) the relative impulse response is represented as a linear time-invariant (LTI) system with the measured input $x(n)$, the measured output $y(n)$, and additive noise $v(n)$. However, $v(n)$ depends on both $x(n)$ and $h(n)$, and thus, standard system identification methods cannot be applied to obtain $h(n)$.

The signals are analyzed in the short-time Fourier transform (STFT) domain using the multiplicative transfer function (MTF) approximation for modeling an LTI system in the STFT domain [13]. Using (2), the cross power spectral density (PSD) between $y(n)$ and $x(n)$ can be written as

$$\lambda_{yx}(l,k) = h_k \lambda_{xx}(l,k) + \lambda_{vx}(k) \quad (3)$$

where l and k represent the time frame and frequency bin indices, respectively, and h_k is the relative transfer function. Replacing the cross-PSD terms with their estimates yields

$$\widehat{\lambda}_{yx}(l,k) = h_k \widehat{\lambda}_{xx}(l,k) + \lambda_{vx}(k) + \varepsilon(l,k) \quad (4)$$

where $\varepsilon(l,k) = \widehat{\lambda}_{vx}(l,k) - \lambda_{vx}(k)$ is the cross-PSD estimation error of $\lambda_{vx}(k)$. Assume N_f time frames of measurements are available. Writing (4) for each time frame $l = 1, \ldots, N_f$ yields N_f distinct equations due to the the non-stationarity of the speech signal. Thus, we obtain a system of N_f linear equations in two variables h_k and $\lambda_{vx}(k)$. The corresponding weighted least square (WLS) estimate of the two variables $\widehat{\boldsymbol{\theta}}_k = [\widehat{h}_k, \widehat{\lambda}_{vx}(k)]^T$ for each frequency bin k is given by

$$\widehat{\boldsymbol{\theta}}_k = \arg\min_{\boldsymbol{\theta}_k} [(\mathbf{s}_{yx}(k) - \mathbf{S}_{xx}(k)\boldsymbol{\theta}_k)^H \times \boldsymbol{\Sigma}_k (\mathbf{s}_{yx}(k) - \mathbf{S}_{xx}(k)\boldsymbol{\theta}_k)] \quad (5)$$

where $\boldsymbol{\Sigma}_k$ is the weight matrix, $(\cdot)^H$ is a conjugate transpose,

$$\mathbf{S}_{xx}(k) = \begin{bmatrix} \widehat{\lambda}_{xx}(1,k) & \cdots & \widehat{\lambda}_{xx}(N_f,k) \\ 1 & \cdots & 1 \end{bmatrix}^T,$$

and $\mathbf{s}_{yx}(k) = [\widehat{\lambda}_{yx}(1,k), \ldots, \widehat{\lambda}_{yx}(N_f,k)]^T$. The solution of (5) with optimal weights that minimize the variance of the estimator [14] is given by

$$\widehat{h}_k = \frac{\langle \widehat{\lambda}_{xx}(l,k)\widehat{\lambda}_{yx}(l,k)\rangle_l - \langle \widehat{\lambda}_{xx}(l,k)\rangle_l \langle \widehat{\lambda}_{yx}(l,k)\rangle_l}{\langle \widehat{\lambda}^2_{xx}(l,k)\rangle_l - \langle \widehat{\lambda}_{xx}(l,k)\rangle_l^2} \quad (6)$$

where $\langle \lambda(l,k)\rangle_l = \sum_{l=1}^{N_f} \lambda(l,k)/N_f$ is an average operator.

3. MANIFOLD OF RTFS

Let $\mathcal{R}$ be a set of RTFs from the region of possible source positions. Based on measurements in controlled noiseless conditions, the RTFs can be estimated using a standard system identification procedure, i.e., $\bar{h}_k = \langle \widehat{\lambda}_{yx}(l,k)/\widehat{\lambda}_{xx}(l,k)\rangle_l$, where $\bar{\mathbf{h}}$ is a vector consisting of all frequency bins values. By collecting all the identified RTFs $\{\bar{\mathbf{h}}^n\}$ we obtain the training set $\mathcal{R}$. We learn the manifold of the training set of the RTFs using an extendable kernel method [10], which relies on Laplacian Eigenmaps [11] and Diffusion Maps [12]. Let $\mathbf{W}^{\mathcal{R}}$ be a kernel defined on the training RTFs, whose (n,m)-th element is given by

$$W^{\mathcal{R}}_{nm} = \exp\left\{-\frac{\|\bar{\mathbf{h}}^n - \bar{\mathbf{h}}^m\|^2}{2\varepsilon}\right\} \quad (7)$$

where $\bar{\mathbf{h}}^n, \bar{\mathbf{h}}^m \in \mathcal{R}$ and $\varepsilon > 0$ is the kernel scale. Setting the kernel scale exceeds the scope of this paper and was studied, for example, in a paper by Hein and Audibert [15].

The eigenvalue decomposition (EVD) of the kernel captures its significant components and may provide a compact parameterization of the manifold of the RTFs. Let N_t be the number of RTFs in the training set $\mathcal{R}$. Thus, the size of the kernel $\mathbf{W}^{\mathcal{R}}$ is $N_t \times N_t$ and the length of each eigenvector is N_t. Moreover, the eigenvectors can be viewed as functions of the training RTFs, where the n-th coordinate of each eigenvector is associated with the n-th training RTF. These functions can describe the data in terms of their natural parameters representing physical meanings. For example, it was shown that the eigenvectors can represent the poles of an autoregressive system [16] as well as the acoustic parameters of RIRs, such as the position of the source [9] or the room reverberation time [17]. Let $\{\mu_j, \boldsymbol{\varphi}_j\}$ be the set of the eigenvalues and eigenvectors of $\mathbf{W}^{\mathcal{R}}$, respectively. It can be shown that the eigenvalues are real and positive, and hence, can be written in a descending order $\mu_0 \geqslant \mu_1 \geqslant \ldots > 0$.

The eigenvectors form a complete basis for any real function on the data. In particular, let i_k be a function that retrieves the k-th frequency bin from each RTF, i.e., $i_k(\bar{\mathbf{h}}^n) = \bar{h}^n_k$. Thus, each coordinate of the training RTFs can be expanded as follows

$$\bar{h}^n_k = \sum_{j=0}^{N_t-1} c_{k,j}\varphi_j(n) \quad (8)$$

where $c_{k,j}$ are the projection coefficients on the basis, i.e., $c_{k,j} = \langle i_k, \boldsymbol{\varphi}_j\rangle \triangleq [\bar{h}^1_k, \ldots, \bar{h}^{N_t}_k]\boldsymbol{\varphi}_j$. Typically, the spectrum of the kernel is fast decaying. According to the decay of the spectrum (the eigenvalues), we determine the dimension ℓ of the manifold and assume that the ℓ eigenvectors associated with the ℓ largest eigenvalues provide accurate parameterization of the manifold. Thus, in (8) we may merely sum over these ℓ eigenvectors.

Next, we utilize the parameterization of the training set in order to get a description of any RTF (not necessary in the training set) from the learned region in the room. Let $\mathbf{A}$ be a non-symmetric kernel defined between any RTF $\mathbf{h}^i$ and each of the RTFs in the training set, whose (i,n)-th element is given by

$$A_{in} = \frac{1}{\omega_n d_i}\alpha(\mathbf{h}^i, \bar{\mathbf{h}}^n), \;\; \bar{\mathbf{h}}^n \in \mathcal{R} \tag{9}$$

where $\alpha(\mathbf{h}^i, \bar{\mathbf{h}}^n) = \exp\left\{-\|\mathbf{h}^i - \bar{\mathbf{h}}^n\|^2/\varepsilon\right\}$, $d_i = \sum_n \alpha(\mathbf{h}^i, \bar{\mathbf{h}}^n)$, and $\omega_n = \sum_i \alpha(\mathbf{h}^i, \bar{\mathbf{h}}^n)/d_i$. In [10], Kushnir et al. showed that the training kernel satisfies $\mathbf{W}^{\mathcal{R}} = \mathbf{A}^T\mathbf{A}$. In addition, the dual kernel $\mathbf{W} = \mathbf{A}\mathbf{A}^T$ can be viewed as an extended kernel, whose (i,j)-th element measures the probability that any two RTFs $\mathbf{h}^i$ and $\mathbf{h}^j$ are associated with the same training RTF [18], and its eigenvectors provide an extended parameterization for any RTF. The construction of the kernels implies: (1) $\mathbf{W}^{\mathcal{R}}$ and $\mathbf{W}$ share the same eigenvalues $\{\mu_j\}$ which are the square of the singular values of $\mathbf{A}$, (2) the eigenvectors $\{\boldsymbol{\varphi}_j\}$ of $\mathbf{W}^{\mathcal{R}}$ are the right singular vectors of $\mathbf{A}$, and (3) the eigenvectors $\{\boldsymbol{\psi}_j\}$ of $\mathbf{W}$ are the left singular vectors of $\mathbf{A}$. The singular value decomposition (SVD) of $\mathbf{A}$ describes the algebraic relation between the eigenvectors, i.e.,

$$\boldsymbol{\psi}_j = \frac{1}{\sqrt{\mu_j}}\mathbf{A}\boldsymbol{\varphi}_j. \tag{10}$$

The aforementioned relationship enables to efficiently extend the eigenvectors to new RTFs without applying the computationally expensive EVD. The extension algorithm consists of two stages. In a training stage, the kernel $\mathbf{W}^{\mathcal{R}}$ is directly calculated based on the training set, and its eigenvalue decomposition is computed. The eigenvectors of the kernel form a learned model for the training set. In the test stage, as new estimates of RTFs become available, we construct $\mathbf{A}$ according to (9), and then compute the extended eigenvectors of $\mathbf{W}$ by exploiting the algebraic relationship given by the SVD in (10). Once the extended parameterization is obtained, we can expand any RTF from the region of interest similarly to (8) as

$$h_k^i = \sum_{j=0}^{\ell-1} c_{k,j}\psi_j(i) + \epsilon_k^i \tag{11}$$

where ϵ_k^i is an error term, which stems from the use of the coefficients $c_{k,j}$ (8) based on training and becomes smaller as the number of training RTFs increases. Substituting $\psi_j(i)$ from (10) into (11) yields

$$h_k^i = \sum_{j=0}^{\ell-1} \mu_j^{-1/2} c_{k,j} \sum_{n=1}^{N_t} A_{in}\varphi_j(n) + \epsilon_k^i. \tag{12}$$

By reordering (12) we get

$$h_k^i = \sum_{n=1}^{N_t} A_{in}D_{nk} = (\mathbf{AD})_{ik} + \epsilon_k^i \tag{13}$$

where $\mathbf{D}$ is matrix that can be computed in advance and whose (n,k)-th element is $D_{nk} = \sum_{j=0}^{\ell-1}\mu_j^{-1/2}c_{k,j}\varphi_j(n)$. When we are interested in finding the parameterization of a single new RTF $\mathbf{h}$, the matrix $\mathbf{A}$ is reduced to a vector, and (13) is rewritten by concatenating all frequency bins as

$$\mathbf{h} = \mathbf{D}^T\mathbf{a} + \boldsymbol{\epsilon} \tag{14}$$

where $\mathbf{a}$ is a vector of length N_t whose n-th element is defined similarly to (9) as $a_n = \exp\left\{-\|\mathbf{h} - \bar{\mathbf{h}}^n\|^2/\varepsilon\right\}$.

4. SUPERVISED RTF IDENTIFICATION

The geometric information extracted from the training set is summarized in (14). It implies that every RTF from the learned region of interest can be expanded by the learned building blocks (the extended eigenvectors). Thus, combining (5) and (14) yields the following constrained identification procedure that relies on both the a priori learned geometric structure and the measured signals

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_k = \arg\min_{\boldsymbol{\theta}_k}[&(\mathbf{s}_{yx}(k) - \mathbf{S}_{xx}(k)\boldsymbol{\theta}_k)^H \\ &\times \boldsymbol{\Sigma}_k(\mathbf{s}_{yx}(k) - \mathbf{S}_{xx}(k)\boldsymbol{\theta}_k)], \;\; \forall k \\ \text{subject to } &\mathbf{h} - \mathbf{D}^T\mathbf{a} \leqslant \xi\end{aligned} \tag{15}$$

where $\xi > 0$ is a small constant. The nonlinear form of the coefficients $\mathbf{a} = \mathbf{a}(\mathbf{h})$ makes (15) hard to solve. Thus, we relax the problem and split it into two stages. In the first stage, we obtain a solution $\widehat{\mathbf{h}}_{wls}$ by solving the WLS problem (6) for each frequency bin, which relies merely on the noisy measurements. In the second stage, we exploit the prior geometric information and project the WLS solution onto the building blocks of the learned manifold, i.e.,

$$\widehat{\mathbf{h}} = \mathbf{D}^T\mathbf{a}(\widehat{\mathbf{h}}_{wls}). \tag{16}$$

Thus, the new estimate is restricted to the learned *manifold* of RTFs (although it is not limited to be one of the training RTFs). Algorithm 1 summaries the identification procedure.

5. EXPERIMENTAL RESULTS

Based on the image method [19], we simulate acoustic impulse responses in a room of dimensions $6 \times 6 \times 3$ m with a moderate reverberation time of $T_{60} = 0.15$ s. We place the two microphones at $(3.1, 1, 1)$ m and $(3.2, 1, 1)$ m. A grid of $N_t = 150$ source positions in a sector of an approximate size of 50×50 cm at a distance of 3 m from the microphones is used for training. Figure 1 illustrates the room setup.

The training data consist of repeated recordings of a speech signal sampled at 16 kHz and duration of 3 s from each position on the grid in noiseless conditions; the anechoic speech signal is convolved with the corresponding simulated impulse responses between the positions on the grid and the

Algorithm 1

Learning the Manifold of RTFs (Training Stage):

1. Obtain training recordings from the region of interest in the room in noiseless conditions.
2. Compute a training set $\mathcal{R}$ of typical RTFs.
3. Construct the kernel $\mathbf{W}^{\mathcal{R}}$ (7).
4. Compute the eigenvalue decomposition $\{\mu_j, \boldsymbol{\varphi}_j\}_j$ of $\mathbf{W}^{\mathcal{R}}$ and compute the projection coefficients $\mathbf{D}$.

Supervised RTF Identification (Test Stage):

1. Obtain a new segment of measurements.
2. Compute the WLS solution of the RTF (6).
3. Confine the RTF to the learned manifold (14).

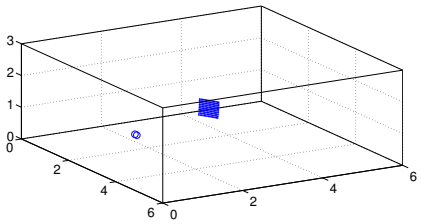

Fig. 1. An illustration of room setup. The circles mark the pair of microphones and the dots mark a grid of possible source positions.

two microphones to generate the microphone measurements. The test data consist of a recorded speech signal, which is generated by convolving the anechoic speech (different from the signal used for the training data) with a simulated impulse response from a random position within the learned sector and the two microphones. In addition, white noise is added to the measured signal. Various signal to noise ratios (SNRs) are examined. The PSD is implemented using relatively long time frames of length 8000 samples to correspond to the long length of the RTFs. For the supervised identification, we use $\ell = 20$ eigenvectors out of the available 150.

Figure 2 illustrates the parameterization of the manifold of impulse responses. We show a scatter plot of the 150 components of the principal eigenvector φ_1 of the training kernel and the x-coordinates of the 150 training positions on the grid. We observe a close to linear correspondence; it implies that the manifold of RTFs has a physical meaning and that the parameterization indeed captures one of the acoustic parameters, which is in this case the position of the source.

The presented supervised RTF identification can be used to build supervised beamformers. For example, the estimated RTFs can be incorporated into the implementation of a blocking matrix in generalized sidelobe canceler (GSC) techniques

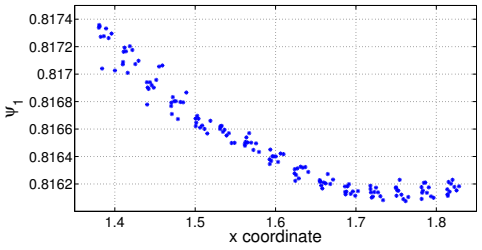

Fig. 2. A scatter plot of the values of the principal eigenvector φ_1 of the training kernel and the x-coordinates of the training positions on the grid.

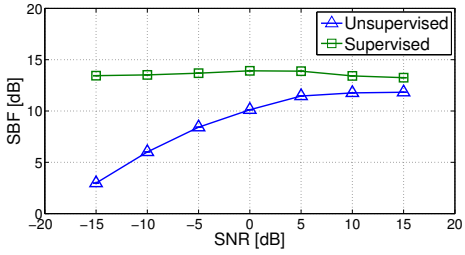

Fig. 3. The SBF curves obtained by the unsupervised nonstationary estimator (blue) and by the proposed supervised RTF estimator (green) as functions of the SNR.

[1, 2]. Thus, we examine the benefit of the supervised method in terms of the blocking ability and use the signal blocking factor (SBF) as an objective quality measure. The SBF is defined as

$$\text{SBF} = 10 \log_{10} \frac{\text{var}(x(n))}{\text{var}(r(n))}$$

where $r(n)$ is the leakage signal defined as $r(n) = h(n) * x(n) - \hat{h}(n) * x(n)$. This measure indicates the ability to block the desired signal and produce reference noise signals. The described experiment is repeated several times with different source positions and different noise realizations and the reported SBF results are the mean values.

Figure 3 depicts the obtained SBF curves as a function of the SNR. We compare the proposed supervised RTF estimator to the unsupervised nonstationary estimator [14], which is given in (6). We observe that the supervised RTF identification achieves higher SBF compared to the unsupervised RTF identification. The obtained improvement is greater in low SNR conditions. In low SNR conditions, the measured signal is less reliable, and hence, the a priori learned model becomes more significant and constraining the identified RTF to the manifold has a greater impact. In addition, relying on the learned manifold makes the supervised RTF identification robust to measurement noise. However, the learned model is limited and the SBF does not increase in SNR higher than 5 dB. Nevertheless, it is still beneficial and yields superior identification compared to the unsupervised identification.

We remark that the supervised RTF estimation was also compared to a simple nearest neighbors search among the training responses and yielded improved results. This demonstrates the significance of *learning the manifold* of RTFs rather than merely using the pre-acquired ones.

6. CONCLUSIONS

We have presented a supervised RTF identification method, in which the manifold of typical RTFs in a particular room is learned in advance, and then, exploited to improve the identification of unknown RTFs based on noisy measurements. Experimental results show that the presented supervised identification method exhibits a superior blocking ability over a competing unsupervised method, especially in noisy conditions. Thus, in a future work, we intend to incorporate this approach into a GSC beamformer. We expect to attain better performance due to the better blocking ability as well as the more accurate construction of the fixed beamformer. In addition, we plan to examine the performance of this method on real recordings in different acoustic and noise conditions. It would be also interesting to explore the effect of environmental changes taking place after the training stage on the identification.

7. REFERENCES

[1] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.

[2] R. Talmon, I. Cohen, and S. Gannot, "Convolutive transfer function generalized sidelobe canceler," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 7, pp. 1420–1434, Sep. 2009.

[3] I. Cohen, "Relative transfer function identification using speech signals," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 5, pp. 451–459, Sep. 2004.

[4] R. Talmon, I. Cohen, and S. Gannot, "Relative transfer function identification using convolutive transfer function approximation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 4, pp. 546–555, May 2009.

[5] M. Fozunbal, T. Kalker, and R.W. Schafer, "Multichannel echo control by model learning," *Proc. IEEE Internat. Workshop Acoust. Echo and Noise Control*, 2008.

[6] T. Koren, R. Talmon, and I. Cohen, "Supervised system identification based on local PCA models," *Proc. IEEE Internat. Conf. Acoust. Speech and Signal Process.*, 2012.

[7] Y. R. Zheng, R. A. Goubran, and M. El-Tanany, "Robust near-field adaptive beamforming with distance discrimination," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 5, pp. 478–488, 2004.

[8] Z. Koldovsky, J. Malek, P. Tichavsky, and F. Nesta, "Semi-blind noise extraction using partially known position of the target source," *to appear in IEEE Trans. Audio, Speech, Lang. Process.*, 2013.

[9] R. Talmon, I. Cohen, and S. Gannot, "Supervised source localization using diffusion kernels," *Proc. IEEE Workshop on Applications of Signal Process. to Audio and Acoust.*, 2011.

[10] D. Kushnir, A. Haddad, and R. Coifman, "Anisotropic diffusion on sub-manifolds with application to earth structure classification," *Appl. Comput. Harm. Anal.*, vol. 32, no. 2, pp. 280–294, 2012.

[11] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, pp. 1373–1396, 2003.

[12] R. Coifman and S. Lafon, "Diffusion maps," *Appl. Comput. Harm. Anal.*, vol. 21, pp. 5–30, Jul. 2006.

[13] Y. Avargel and I. Cohen, "On multiplicative transfer function approximation in the short time Fourier transform domain," *IEEE Signal Process. Let.*, vol. 14, pp. 337–340, 2007.

[14] O. Shalvi and E. Weinstein, "System identification using nonstationary signals," *IEEE Trans. Signal Process.*, vol. 40, no. 8, pp. 2055–2063, Aug. 1996.

[15] M. Hein and J. Y. Audibert, "Intrinsic dimensionality estimation of submanifold in r^d," *L. De Raedt, S. Wrobel (Eds.), Proc. 22nd Int. Conf. Machine Learning, ACM*, pp. 289–296, 2005.

[16] R. Talmon, D. Kushnir, R. Coifman, I. Cohen, and S. Gannot, "Parametrization of linear systems using diffusion kernels," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1159–1173, Mar. 2012.

[17] R. Talmon and E. A. P. Habets, "Blind reverberation time estimation by intrinsic modeling of reverberant speech," *Proc. IEEE Internat. Conf. Acoust. Speech and Signal Process.*, May 2013.

[18] R. Talmon, I. Cohen, S. Gannot, and R. Coifman, "Supervised graph-based processing for sequential transient interference suppression," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 20, no. 9, pp. 2528–2538, 2012.

[19] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am*, vol. 65, no. 4, pp. 943–950, 1979.