EUSIPCO 2013 1569742725

GRAPH-BASED BAYESIAN APPROACH FOR TRANSIENT INTERFERENCE SUPPRESSION

Ronen Talmon[1], *Israel Cohen*[2], *Sharon Gannot*[3], *and Ronald R. Coifman*[1]

[1]Department of Mathematics, Yale University
[2]Department of Electrical Engineering, Technion - Israel Institute of Technology
[3]Faculty of Engineering, Bar-Ilan University

ABSTRACT

In this paper, we present a method for transient interference suppression. The main idea is to learn the intrinsic geometric structure of the transients instead of relying on estimates of noise statistics. The transient interference structure is captured via a parametrization of a graph constructed from the measurements. This parametrization is viewed as an empirical model for transients and is used for building a filter that extracts transients from noisy speech. We present a model-based supervised algorithm, in which the graph-based empirical model is constructed in advance from training recordings, and then extended to new incoming measurements. This paper extends previous studies and presents a new Bayesian approach for empirical model extension that takes into account both the structure of the transients as well as the dynamics of speech signals.

Index Terms— Speech enhancement, transient noise, graph filtering, empirical models.

1. INTRODUCTION

Transient interferences, e.g. keyboard typing and door knocking, are examples of signals whose representations based on statistical models are often poor. To date, most of existing noise reduction algorithms are based on spectral estimation of stationary noise from segments in which the desired signal is absent. Clearly, this approach does not suit the abrupt nature of transient noise; hence such algorithms are inadequate for this problem. Recently, we have presented a novel method for clustering and suppression of transient interferences that circumvented this assumption [1, 2, 3]. The key idea was to learn the intrinsic geometric structure of the transients instead of relying on estimates of noise statistics. We utilized manifold learning approaches, e.g. diffusion maps [4], to embed the measurements into a new domain and exploited the new representation to estimate the spectrum of the transients. Such an approach enabled us in [3] to present a model-based supervised algorithm, similarly to recent studies that rely on training and predefined models [5, 6]. However, a known shortcoming of manifold learning methods is that the dynamics is not considered. The input of these methods is usually a data set of samples given in an arbitrary order, and hence, the temporal cue of signals, such as speech and audio, is ignored although it may convey important and critical information. For example, the duration of typical phonemes can circumvent false identification of transients; a time segment has a higher likelihood to contain speech if the preceding segement contains speech.

In this paper, we extend [3] by considering the dynamics of the measurements. The transient interference structure is captured via a parametrization of a graph constructed from the measurements. This parametrization is viewed as an empirical model for transients and is used for building a filter that extracts transients from noisy speech. The graph-based empirical model is constructed in advance from training recordings, and then extended to new incoming measurements. We present a new Bayesian approach for the model extension that takes into account both the structure of the transients as well as the dynamics of speech signals.

This paper is organized as follows. In Section 2, we formulate the problem. In Section 3, the construction of local models for transients is presented. In Section 4, we build the empirical model based on a graph, describe a Bayesian approach for extending the model to new incoming measurements by exploiting the dynamics, and define a filter to estimate the spectrum of the transients. In Section 5, the speech enhancement procedure using the transient spectrum estimate is reviewed. Finally, in Section 6, experimental results are presented, demonstrating the impact of taking the dynamics into account.

2. PROBLEM FORMULATION

Let $x(n)$ denote a clean speech signal measured by a single microphone. The measured signal $y(n)$ is given by

$$y(n) = x(n) + t(n) + u(n) \tag{1}$$

where $t(n)$ and $u(n)$ are additive transient interference and stationary background noise, respectively. The transient component $t(n)$ may consist of multiple types of interferences.

This research was supported by the Israel Science Foundation (grant no. 1130/11). The work of R. Talmon was supported in part by the Viterbi Fellowship, Technion

Rewriting (1) in the short-time spectrum domain, assuming the speech, the transient interference, and the stationary noise are mutually uncorrelated, yields

$$\lambda_y(l,k) = \lambda_x(l,k) + \lambda_t(l,k) + \lambda_u(l,k) \tag{2}$$

where $\lambda_y(l,k)$, $\lambda_x(l,k)$, $\lambda_t(l,k)$, and $\lambda_u(l,k)$ are the short-time power spectral density (PSD) of the signals.

In this work, our focus is on estimating the PSD of the transient interference $\lambda_t(l,k)$ given noisy measurements. Then, the estimated spectrum is incorporated into an existing speech enhancement algorithm to obtain simultaneous suppression of transient and background noise.

3. LOCAL MODELS OF TRANSIENTS

Suppose training sets of typical transient instances $\{\mathcal{T}_i\}_i$ are available in advance, where each set $\mathcal{T}_i$ contains instances of the same transient type. We define an empirical model for each transient type based on these training sets. In order to exploit the spectral structure of the transients, we collect the spectral features from all the frequency bins of each time frame into vectors. Let $\boldsymbol{\lambda}_t(l)$ be a vector of the PSD values, defined by

$$\boldsymbol{\lambda}_t(l) = [\lambda_t(l,0), \ldots, \lambda_t(l,N-1)]^T \tag{3}$$

where N is the number of frequency bins. Each of the vectors can be viewed as an N-dimensional feature vector. Let $\bar{\boldsymbol{\eta}}_i$ be the empirical mean vector of the i-th training set, i.e.,

$$\bar{\boldsymbol{\eta}}_i = \frac{1}{|\mathcal{T}_i|} \sum_{l \in \mathcal{T}_i} \log\left(\boldsymbol{\lambda}_t(l)\right)$$

and let $\mathbf{C}_i$ be the empirical covariance matrix of the set

$$\mathbf{C}_i = \frac{1}{|\mathcal{T}_i|} \sum_{l \in \mathcal{T}_i} \left(\log\left(\boldsymbol{\lambda}_t(l)\right) - \bar{\boldsymbol{\eta}}_i\right)\left(\log\left(\boldsymbol{\lambda}_t(l)\right) - \bar{\boldsymbol{\eta}}_i\right)^T$$

where $|\mathcal{T}_i|$ is the cardinality of the set $\mathcal{T}_i$. The pair $(\bar{\boldsymbol{\eta}}_i, \mathbf{C}_i)$ may be used as the learned model of the i-th transient type. We choose the logarithmic domain over the linear domain since empirical experiments show better results.

A well-known limitation of principal component analysis (PCA) is that it is linear and able to capture only the global structure of the training data. Thus, in this work we apply PCA locally to each transient interference type. We obtain that the eigenvectors of $\mathbf{C}_i$, associated with the largest eigenvalues, capture most of the information disclosed in each set, and hence, we may use only the subspace spanned by a few principal eigenvectors. Let $\{\mathbf{v}_{i,j}\}_{j=1}^{L}$ be the set of L such principal eigenvectors.

Define P_i to be a linear projection operator of any spectral feature vector onto the local model of the i-th transient type, given by

$$P_i(\boldsymbol{\lambda}(l)) = \bar{\boldsymbol{\eta}}_i + \sum_{j=1}^{L} \langle \log\left(\boldsymbol{\lambda}(l)\right) - \bar{\boldsymbol{\eta}}_i, \mathbf{v}_{i,j} \rangle \mathbf{v}_{i,j}. \tag{4}$$

Based on this projection, we define a pairwise metric between spectral feature vectors as

$$d_i\left(\boldsymbol{\lambda}(l), \boldsymbol{\lambda}(l')\right) = \left\| P_i(\boldsymbol{\lambda}(l)) - P_i(\boldsymbol{\lambda}(l')) \right\|. \tag{5}$$

The projection in (4) extracts the components of the i-th transient type from the feature vector of any measurement, and hence, the metric in (5) compares measurements only in terms of the i-th type.

4. TRANSIENT SPECTRUM ESTIMATION

4.1. Graph-Based Global Model

Let $\mathcal{T}$ denote the collection of all training sets. We define a non symmetric kernel $\mathbf{A}$ consisting of an affinity measure between the feature vector of any measurement $\boldsymbol{\lambda}_y(l)$ (not included in the training set) and the feature vector of any training sample $\boldsymbol{\lambda}_t(p)$, where $\boldsymbol{\lambda}_y(l)$ is defined similarly to (3). The (l,p)-th element of the kernel is given by

$$A^{lp} = \frac{1}{\omega_l} \exp\left\{ -\frac{d_i\left(\boldsymbol{\lambda}_y(l), \boldsymbol{\lambda}_t(p)\right)}{\sigma^2} \right\} \tag{6}$$

where $p \in \mathcal{T}_i$, σ^2 is the kernel scale, and ω_l is a normalization factor that satisfies $\sum_{p \in \mathcal{T}} A^{lp} = 1$ for all l.

We now define two symmetric kernels: $\mathbf{W}_t = \mathbf{A}^T\mathbf{A}$ and $\mathbf{W} = \mathbf{A}\mathbf{A}^T$. The kernel $\mathbf{W}_t$ is defined on the training set and consists of an affinity between any pair of training samples. On the other hand, the kernel $\mathbf{W}$ consists of an affinity between any pair of measurements with respect to the training set. It further implies that two measurements are similar if they "see" the training samples in the same way [7]. The two kernels can be viewed as graphs, where the samples are the nodes of the graphs and the kernels determine the weights of the edges connecting the nodes. For example, nodes $\boldsymbol{\lambda}_y(l)$ and $\boldsymbol{\lambda}_y(l')$ are connected by an edge with weight $W^{ll'}$. The role of the graph is to integrate the relations between the local models into a global intrinsic model [8, 4]. For more details and a probabilistic interpretation see [3].

We proceed by drawing the algebraic connection between the eigen-decomposition of the kernels $\mathbf{W}_t$ and $\mathbf{W}$. The eigenvectors of the kernels are parametrization of the underlying structures of the samples and viewed as empirical models [8, 4]. Let $\{\mu_j, \boldsymbol{\varphi}_j, \boldsymbol{\psi}_j\}_j$ be the singular value decomposition (SVD) of $\mathbf{A}$. 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 construction of the kernels $\mathbf{W}_t$ and $\mathbf{W}$ implies that they share the same eigenvalues μ_j^2, and $\boldsymbol{\varphi}_j$ and $\boldsymbol{\psi}_j$ are the eigenvectors of $\mathbf{W}_t$ and $\mathbf{W}$, respectively. The aforementioned relationship is especially suitable for sequential extension of the eigenvectors to new incoming measurements consisting of two stages. In a training stage, the kernel $\mathbf{W}_t$ is directly computed based on the training sets, and its eigen-decomposition is calculated. The eigenvectors of the

kernel form a learned model for the training set. In a test stage, as new incoming measurements become available, we construct $\mathbf{A}$ according to (6), and then compute the extended eigenvectors of $\mathbf{W}$ by exploiting the algebraic relationship given by the SVD of $\mathbf{A}$

$$\boldsymbol{\psi}_i = \frac{1}{\mu_i}\mathbf{A}\boldsymbol{\varphi}_i. \tag{7}$$

It is worthwhile noting that the extension does not involve an eigen-decomposition. Thus, the processing of new measurements requires low computational complexity (see [3]).

4.2. Bayesian Approach

We present a nonparametric Bayesian framework [9] that incorporates the dynamics of the measurements to improve the extension of the empirical model. In terms of a standard Bayesian derivation, we view the estimate of the extended eigenvector in (7) as the available noisy "measurement", and we seek the "state", which is the true extended eigenvector $\boldsymbol{\psi}_j^*$. We assume that the likelihood function of the "measurement" given the "state" is locally approximated for each eigenvector j by a normal distribution, i.e.,

$$\psi_j(l)|\psi_j^*(l) \sim \mathcal{N}(\psi_j^*(l), \boldsymbol{\Sigma}_{j,l}) \tag{8}$$

where $\boldsymbol{\Sigma}_{j,l}$ is the local covariance near $\boldsymbol{\psi}_j^*(l)$.

We proceed by incorporating the empirical dynamics of past observations. We note that the l-th coordinate of each eigenvector parameterizes the PSD vector in the l-th time frame $\boldsymbol{\lambda}_y(l)$. Let $\mathcal{N}_{l-1}$ be a set of samples in a $\xi > 0$ neighborhood of $\widehat{\boldsymbol{\lambda}}_x(l-1)$, defined as

$$\mathcal{N}_{l-1} = \left\{ p \left| \left\| \widehat{\boldsymbol{\lambda}}_x(p) - \widehat{\boldsymbol{\lambda}}_x(l-1) \right\| < \xi, \;\; p < l-1 \right. \right\}$$

where $\widehat{\boldsymbol{\lambda}}_x(l-1)$ is the estimate of the spectrum of the speech in the preceding time frame, which is the output of the algorithm. The samples in this neighborhood represent similar past states and can be used for dynamics estimation since their succeeding samples are available. We examined several ways to convey the dynamics and empirically concluded that searching similar states based on the speech estimate yields maximal performance. Since the transient interference has an abrupt nature and the background noise is stationary, the speech is the primary component in the measurement with dynamics. We collect the succeeding samples (of the true "state" obtained in past steps), i.e., $\psi_j^*(p+1)$ for each $p \in \mathcal{N}_{l-1}$, and compute their mean and covariance, denoted by $\bar{\psi}_{j,l-1}^f$ and $\boldsymbol{\Sigma}_{j,l-1}^f$, respectively. The pdf of the dynamics is estimated by a normal distribution and given by

$$\psi_j^*(l)|\psi_j^*(l-1) = \mathcal{N}(\bar{\psi}_{j,l-1}^f, \boldsymbol{\Sigma}_{j,l-1}^f). \tag{9}$$

Since we merely have pointwise definitions of the statistical models, we use the concept of sequential Monte Carlo methods [10] and represent the posterior pdf by a set of support samples $\{\psi_j^{(k)}(l)\}_{k=1}^P$ ("particles"), i.e.,

$$\begin{aligned} &\Pr(\psi_j^*(l)|\psi_j^*(l-1), \psi_j(l)) \\ &\quad\approx \sum_{k=1}^{P} w_l^{(k)} \delta\left(\psi_j^*(l) - \psi_j^{(k)}(l)\right), \end{aligned} \tag{10}$$

where the weights are given by

$$w_l^{(k)} = \Pr(\psi_j^{(k)}(l)|\psi_j^*(l-1), \psi_j(l)),$$

with $\sum_{k=1}^P w_t^{(k)} = 1$. We note that in (10) we assume Markovian dynamics, i.e., that the current state depends only on the previous state. We therefore have a discrete weighted approximation of the desired posterior pdf. At each time step, the particles are drawn from the posterior pdf estimate of the preceding step. By Bayes' theorem and by the Markov dynamical model, we obtain that

$$w_l^{(k)} \propto \Pr(\psi_j^{(k)}(l)|\psi_j^*(l-1)) \Pr(\psi_j(l)|\psi_j^{(k)}(l)). \tag{11}$$

The estimates of the densities in (11) are given by (8) and (9). Using the estimate of the posterior pdf, a sequential estimator can be computed according to an optimization criterion. For example, using (10) the minimum mean squared error (MMSE) estimator is given by

$$\begin{aligned} \widehat{\psi}_j^*(l) &= \mathbb{E}\left[\psi_j^*(l)|\psi_j^*(l-1), \psi_j(l)\right] \\ &\approx \sum_{k=1}^{P} w_t^{(k)} \psi_j^{(k)}(l). \end{aligned} \tag{12}$$

After obtaining all new estimates we apply the Gram-Schmidt process to maintain orthogonality.

4.3. Graph-based Filter

By the orthogonality of the eigenvectors, the set $\{\boldsymbol{\psi}_j^*\}_j$ forms a complete basis for any real function defined on the set of PSD vectors $\{\boldsymbol{\lambda}_y(l)\}_l$. In particular, let i_k be a function that retrieves the k-th frequency bin from the PSD vector $\boldsymbol{\lambda}_y(l)$, i.e., $i_k\left(\boldsymbol{\lambda}_y(l)\right) = \lambda_y(l,k)$. It implies that each spectral component can be expanded by the set of eigenvectors as

$$\lambda_y(l,k) = i_k\left(\boldsymbol{\lambda}_y(l)\right) = \sum_j \mu_j^2 \langle i_k, \boldsymbol{\psi}_j^* \rangle \boldsymbol{\psi}_j^*(l)$$

where the inner product is defined as

$$\langle i_k, \boldsymbol{\psi}_j^* \rangle \triangleq [\ldots, \lambda_y(\cdot,k), \ldots]\, \boldsymbol{\psi}_j^*.$$

The parametrization of the graph captures the dominant structures of the measurements. Since the construction of the graph with an affinity metric based on local transient models

was designed to emphasize transients, we assume that there exists a subset of ℓ eigenvectors which represent the transient interference. For simplicity, we assume that this subset consists of the eigenvectors associated with the largest eigenvalues, i.e., $\{\boldsymbol{\psi}_j^*\}_{j=0}^{\ell-1}$. In practice, we may determine the appropriate eigenvectors by observing their spectral structure.

We define the following graph-based filter that estimates the transient PSD by projecting the PSD of the measurements onto the eigenvectors spanning the transient interference subspace

$$\widehat{\lambda}_t(l,k) = \sum_{j=0}^{\ell-1} \mu_j^2 \langle i_k, \boldsymbol{\psi}_j^* \rangle \boldsymbol{\psi}_j^*(l). \tag{13}$$

In practice, few speech "leftovers" may appear in the PSD estimate. Human speech consists of both harmonic and non-harmonic sounds and it can span across a wide range of frequencies. Thus, many speech phonemes can be represented (at least partially) by the transients "building blocks". Such residuals in the PSD estimate of the transient signal degrade the quality of the speech when incorporated into an enhancement algorithm. Since the leftovers usually exist in periods where the transient signal is absent, we are able to easily identify them by their low magnitude compared to the magnitude of the transients. Thus, we remove potential leftovers by employing a hard threshold.

5. SPEECH ENHANCEMENT

We employ the optimally modified log spectral amplitude (OM-LSA) estimator [11], in which the optimal spectral gain with respect to the minimum log spectral amplitude (LSA) error criterion is controlled by the speech presence probability. Since it is unknown, the speech presence probability is estimated based on the time-frequency distribution of the a-priori signal-to-noise ratio (SNR), where the noise variance is estimated using the improved minima controlled recursive averaging (IMCRA) [12]. Unfortunately, short and abrupt bursts of transient interferences are falsely detected as speech components. Hence, the transient interference is not included in the noise PSD estimate obtained by the IMCRA approach, and as a result, is not attenuated. In this work, we set the optimal spectral gain to correspond to the sum of the PSD estimates of the transient interference $\widehat{\lambda}_t(l,k)$ and the stationary noise $\widehat{\lambda}_u(l,k)$. The former estimate is obtained by the graph-based filter (13) followed by the hard thresholding, and the latter estimate is obtained by the IMCRA. The IMCRA and the OM-LSA parameters used in this stage are similar to the set of parameters used to enhance speech and reduce stationary background noise as described in [11]. Since the optimal spectral gain is controlled by the transient interference spectrum, the suppression of transients is now attainable.

6. EXPERIMENTAL RESULTS

We evaluate the performance of the proposed algorithm and compare it to the method presented in [3] to demonstrate the impact of the Bayesian approach. Similarly to the experimental setup in [3] we use recorded speech and transient signals sampled at 16 KHz. The time domain measurements are constructed according to (1). We rescale the speech and transient interference to have equal maximal amplitude in the measured interval. The additive stationary noise part is a white Gaussian noise with SNR of 20 dB. Each measurement is 20 s long and comprises several speech utterances of 5 different speakers and 30 transient events. For the time-frequency representation, we use time frames of 512 samples length with 75% overlap between successive frames. We have examined the suppression of three transient interference signals: (1) keyboard typing, (2) household interferences, and (3) door knocks. We trained three corresponding models based on training recordings consisting of 10 instances of transients from each type. In addition, in order to represent the transients and define the graph-based filter (13) we used the principal $\ell = 10$ eigenvectors of the graph. For each transient interference we empirically tuned the parameters, e.g. the kernel scale, such that maximum performance was obtained.

The algorithms are evaluated using three objective measures: SNR, mean log spectral distortion (LSD), and perceptual evaluation of speech quality (PESQ). The SNR and LSD are computed only in time periods where the estimate of the PSD of transients exists. This way we are able to focus on the performance of the proposed algorithm and to evaluate the speech enhancement and the artifacts introduced by the algorithm simultaneously. In periods where the transient estimate does not exit, only stationary noise suppression is attained, and the performance of the algorithm is equal to the performance of the OM-LSA. The PESQ score, which covers different aspects and complements the former quality measures, is computed in the entire segment. The aforementioned experiment was repeated several times with different speech segments and different transient and stationary noise realizations; the reported results are the averages of the measures over these experiments.

Table 1 summarizes the evaluation of the speech enhancement. In all of the tested cases, the addition of Bayesian filtering demonstrates substantially superior LSD and PESQ improvements at the expense of slightly inferior SNR improvements. We note that the PESQ scores of the noisy signals are 2.165, 2.028 and 1.933, respectively. LSD and PESQ improvements indicate that the enhanced speech exhibits fewer distortions and artifacts using the Bayesian filtering, whereas the SNR degradation implies weaker noise suppression.

One of the advantages of incorporating the temporal dynamics is a decrease in the number of "false alarms", i.e., false identification of speech phonemes as transients. For example, an empirical estimation of the dynamics and cross-relations

Table 1. Speech Enhancement Evaluation.

Transient Type	SNR Improvement [dB]		LSD Improvement [dB]		PESQ Score Improvement	
	Method Proposed in [3]	Bayesian Filtering	Method Proposed in [3]	Bayesian Filtering	Method Proposed in [3]	Bayesian Filtering
Keyboard Typing	7.78	7.63	2.12	2.48	0.749	0.753
Household Knocks	6.62	6.57	2.04	2.83	0.644	0.801
Door Knocks	9.79	9.55	2.39	2.52	0.536	0.699

Table 2. Mean Number of False Alarms.

Transient Type	Method Proposed in [3]	Bayesian Filtering
Keyboard Typing	3.3	3
Household Knocks	5.3	4.7
Door Knocks	4	4

between time frames enables to eliminate false transient identifications within speech segments, since a high power measurement has a higher likelihood to contain speech if the preceding measurement contains speech. We present in Table 2 the mean number of encountered false alarms over the different test recordings in the experiments out of the 30 transient events in each recording. We observe that the addition of Bayesian filtering that incorporates the dynamics indeed reduces the number of false alarms. We also note that no missed hits were encountered.

7. CONCLUSIONS

We have presented a supervised graph-based processing framework for sequential transient interference suppression. The primary focus is on building empirical models for transients driven by examples, which can in turn be extended to new measurements in a sequential manner. This paper extends a previous work and introduces a nonparametric Bayesian approach that exploits the dynamics of the speech to improve the extension of the empirical models. Experimental results show that the addition of the dynamics enhances the performance of transient interference suppression and attains lower speech distortion. We note that the presented nonparametric Bayesian framework extends the scope of this paper and can be used to extend empirical models based on spectral representations in the context of time series analysis and processing.

8. REFERENCES

[1] R. Talmon, I. Cohen, and S. Gannot, "Transient noise reduction using nonlocal diffusion filters," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 6, pp. 1584–1599, Aug. 2011.

[2] R. Talmon, I. Cohen, and S. Gannot, "Single-channel transient interference suppression using diffusion maps," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 21, no. 1, pp. 132–144, Jan. 2013.

[3] R. Talmon, I. Cohen, S. Gannot, and R.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.

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

[5] S. Srinivasan, J. Samuelsson, and W. B. Kleijn, "Codebook-based Bayesian speech enhancement for nonstationary environments," *IEEE Trans. Audio Speech Lang. Process.*, vol. 15, pp. 441–452, 2007.

[6] K.W. Wilson, B. Raj, P. Smaragdis, and A. Divakaran, "Speech denoising using nonnegative matrix factorization with priors," *ICASSP-2008*, pp. 4029 – 4032, 2008.

[7] 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.

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

[9] R. Talmon and R.R. Coifman, "Empirical intrinsic geometry for nonlinear modeling and time series filtering," *in press, Proc. Nat. Acad. Sci.*, 2013.

[10] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, pp. 197–208, 2000.

[11] I. Cohen and B. Berdugo, "Speech enhancement for non stationary noise environments," *Signal Process.*, vol. 81, no. 11, pp. 2403–2418, Nov. 2001.

[12] I. Cohen, "Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 5, pp. 466–475, Sep. 2003.