

BLIND REVERBERATION TIME ESTIMATION BY INTRINSIC MODELING OF REVERBERANT SPEECH

Ronen Talmon

Department of Mathematics
Yale University
New Haven, CT 06520, U.S.A.

Emanuël A. P. Habets

International Audio Laboratories Erlangen
Am Wolfsmantel 33
91058 Erlangen, Germany

ABSTRACT

The reverberation time (RT) is a very important measure that quantifies the acoustic properties of a room and provides information about the quality and intelligibility of speech recorded in that room. Moreover, information about the RT can be used to improve the performance of automatic speech recognition systems and speech dereverberation algorithms. In a recent study, it has been shown that existing methods for blind estimation of the RT are highly sensitive to additive noise. In this paper, a novel method is proposed to blindly estimate the RT based on the decay rate distribution. Firstly, a data-driven representation of the underlying decay rates of several training rooms is obtained via the eigenvalue decomposition of a specially-tailored kernel. Secondly, the representation is extended to a room under test and used to estimate its decay rate (and hence its RT). The presented results show that the proposed method outperforms a competing method and is significantly more robust to noise.

Index Terms— reverberation time, blind estimation, room acoustics

1. INTRODUCTION

The reverberation time (RT) is a very important measure that quantifies the acoustic properties of a room. The RT is defined as the time it takes for the sound to decay by 60 dB once the source has been switched off [1]. The RT highly depends on the room geometry and the reflectivity of the surfaces in the room and is commonly given by the Sabine or the Eyring equations [2]. In contrast to the room impulse response (RIR), the RT is independent of the source-microphone configuration. An estimate of the RT of a room can serve as an indicator of the quality and the intelligibility of speech observed in that room. Moreover, it can be used to improve the performance of automatic speech recognition systems [3, 4] and dereverberation algorithms [5, 6].

Both channel-based and signal-based methods have been developed to estimate the RT. The channel-based methods require an estimate of the RIR. The most commonly used channel-based method calculates the energy decay curve of the RIR using the Schroeder backward integration method [7] and fits a line to its slope in some range (typically from -5 to -35 dB) depending on the estimated noise floor. The RT is then computed based on the slope of the line. Although, this provides accurate estimates of the RT, it may not always be practical or even possible to measure the RIR in a room. Therefore, it is desirable to be able to estimate the RT directly from an observed reverberant speech signal. Several methods have been proposed to blindly estimate the RT [8–14]. In [10], Wen et al. proposed a method that blindly estimates the RT by analyzing the distribution

of decay rates of the observed reverberant speech signal. The authors shown that the negative-side variance of the distribution can be related to the RT. The method requires a training phase to obtain the relation between the negative-side variance and the RT. In a recent study Gaubitch et al. compared three different RT estimation methods [15]. It was shown that the performance of all existing methods significantly decreases when the signal-to-noise ratio decreases.

In this paper, we propose a novel method to blindly estimate the RT based on the decay rate distribution. Instead of using a specific characteristic of the distribution (such as the negative-side variance used in [10]), the proposed method empirically reveals the most significant underlying parameter of the decay rates of the observed reverberant speech signal. It is shown that this parameter is strongly related to the decay rate of the room. Firstly, a data-driven representation of the underlying decay rates of several training rooms is obtained via the eigenvalue decomposition of a kernel. Unlike common kernel methods, this kernel is built based on a specially-tailored distance between the observable decay rate distributions of the reverberant speech and is shown to uncover intrinsic geometric information on the underlying parameter. Secondly, the representation is extended to a room under test and used to estimate its decay rate (and hence its RT). A major advantage of the proposed method is its robustness to additive noise. The presented results show that the RT can be estimated with a root mean squared error smaller than 50 ms for SNRs larger than 0 dB.

2. PROBLEM FORMULATION

In this paper, we assume that each room is characterized by merely a single decay rate value of the energy envelope, which is independent of the frequency. Let λ_r denote the characteristic decay rate of a room r . Under the Polack model for RIRs [16], the decay rate λ_r is related to the RT by [10]

$$\text{RT} = -6 \ln(10) / \lambda_r. \quad (1)$$

In the following, we rely on the fact that there is a one-to-one mapping between the decay rate of a room and the RT. Estimating the decay rate of a room and estimating the RT are therefore considered equivalent tasks, which are used interchangeably in this paper.

Let \mathcal{R} be a collection of training rooms with various known characteristic decay rates. In each room $r \in \mathcal{R}$ with a characteristic decay rate λ_r , we perform L recordings and collect a set of L reverberant speech signals, denoted by $\{x_r^{(i)}(n)\}_{i=1}^L$. Since each recording may be made from a different location in the room and the room itself may be subject to small movements of objects between the different recordings, we assume that the decay rate of the RIR in each

experiment is slightly perturbed. Let $\{\lambda_r^{(i)}\}_{i=1}^L$ be the set of random variables that represent the estimates of the decay rates of the energy envelopes of the RIRs corresponding to the L measurements. Each variable is expressed as

$$\lambda_r^{(i)} = \lambda_r + \epsilon^{(i)} \quad (2)$$

where $\epsilon^{(i)}$ is a Gaussian random variable with zero mean and arbitrary variance which represents the variation of the i -th recording.

Let λ_s be a random variable that represents the instantaneous decay rate of the energy envelope of an anechoic speech. The decay rates of the room and the anechoic speech are unobservable and may be estimated via the measured reverberant speech. Reverberant speech in an enclosure is usually modeled as the convolution of the anechoic speech signal and the RIR. Thus, the energy envelope of the reverberant speech signal can be viewed as a function of the energy envelopes of the anechoic speech signal and the RIR. Let λ_x denote the *observable* instantaneous decay rate of the energy envelope of the reverberant speech, which can be written as

$$\lambda_x = g(\lambda_r, \lambda_s) \quad (3)$$

where g is an arbitrary (possibly nonlinear) function. We note that for simplicity the room index is omitted from λ_x .

Our objective in this paper is to recover the decay rate (RT) of a room from λ_x without model assumptions. We assume that accurate estimates of λ_x can be obtained from the observable reverberant speech signal. The decay rates can be estimated in the time-frequency domain according to [10] or in the time-domain according to [14]. We note that in [10], a model-based analysis was carried out to determine a closed form expression for g . In this paper, we restrain from making any model assumptions and propose a data-driven method.

The general structure of the proposed method is as follows: the training sets of recordings are used for learning a data-driven representation in advance; then, based on the learnt representation, we propose to estimate the decay rate λ_r (RT) of an “unseen” room ($r \notin \mathcal{R}$) from a single reverberant speech measurement.

3. DECAY RATE DISTRIBUTION

Let $f(\lambda_r)$ and $q(\lambda_s)$ denote the probability density functions (pdfs) of λ_r and λ_s , respectively. By (3) and by assuming that λ_r and λ_s are independent, the pdf of λ_x , denoted by $p(\lambda_x)$, is given by

$$p(\lambda_x) = \int_{\lambda_x = g(\lambda_r, \lambda_s)} f(\lambda_r)q(\lambda_s)d\lambda_r d\lambda_s. \quad (4)$$

In practice, we propose to estimate the pdfs of the decay rates of reverberant speech signals using histograms. By assuming (unrealistically) that infinite number of decay rate instances are available and that their density in each histogram bin is uniform, each coordinate of the histogram can be written as an integration of the pdf over the corresponding histogram bin.

The above analysis leads to the following observation. The histograms of the observable decay rates of reverberant speech signals acquired in different recordings in a room is a linear function of $f(\lambda_r)$, which is the pdf of the decay rate of the room. This observation motivates processing in the histogram domain and is exploited in Sec. 4. As a result, we consider the histograms as feature vectors of the data.

In the presence of measurement noise, the decay rate of the noisy reverberant speech can be expressed similarly to (3) as $\lambda_x =$

$g(\lambda_r, \lambda_s, \xi)$, where ξ denotes a noise process. Assuming independent noise and following the derivation in (4), the observation that the histograms of the observable decay rates of the noisy reverberant speech signals is a linear function of the pdf of the decay rate of the rooms still holds. Thus, the histograms provide robust features to noise.

From each training recording $x_r^{(i)}(n)$ we compute the decay rates in short time frames. Let $\mathbf{h}_r^{(i)}$ denote the histogram of the decay rates corresponding to the i -th recording in room r . We compute the empirical covariance matrix of the histograms from the same training room as follows

$$\mathbf{C}_r = \frac{1}{L} \sum_{i=1}^L (\mathbf{h}_r^{(i)} - \bar{\mathbf{h}}_r)(\mathbf{h}_r^{(i)} - \bar{\mathbf{h}}_r)^T \quad (5)$$

where $\bar{\mathbf{h}}_r$ is the empirical mean of the histograms in r , for all $r \in \mathcal{R}$. The natural variations of the decay rates in different recordings (2) introduce variations of the corresponding histograms in the observable domain (histograms of the decay rates of the measured reverberant speech). In Sec. 4, we exploit these variations, as manifested in the covariance matrix \mathbf{C}_r , to empirically invert the function g and reveal the decay rates.

4. INTRINSIC DISTANCE FUNCTION AND MODEL

We define a symmetric distance function between pairs of training feature vectors (histograms) as

$$d_{\mathbf{C}}^2(\mathbf{h}_r^{(i)}, \mathbf{h}_\rho^{(j)}) = (\mathbf{h}_r^{(i)} - \mathbf{h}_\rho^{(j)})^T (\mathbf{C}_r^{-1} + \mathbf{C}_\rho^{-1})(\mathbf{h}_r^{(i)} - \mathbf{h}_\rho^{(j)}) \quad (6)$$

for each $r, \rho \in \mathcal{R}$ for all i, j .

The distance in (6) is termed the Mahalanobis distance and has two important properties [17, 18]. The Mahalanobis distance is invariant to linear transformations. Thus, according to the analysis in Sec. 3, in the features (histograms) domain, this distance is invariant to the distortions imposed on the decay rate by the anechoic speech and noise. In addition, it can be shown that the Mahalanobis distance approximates the Euclidean distance between the decay rates of the room, i.e.,

$$|\lambda_r - \lambda_\rho|^2 = d_{\mathbf{C}}^2(\mathbf{h}_r^{(i)}, \mathbf{h}_\rho^{(j)}) + \mathcal{O}(\|\mathbf{h}_r^{(i)} - \mathbf{h}_\rho^{(j)}\|^4). \quad (7)$$

Given the pairwise distances between the desired values, we recover the values themselves through the eigenvalue decomposition (EVD) of an appropriate Laplace operator [19]. Let $\mathbf{W}_{\mathcal{R}}$ be an affinity matrix (kernel) between pairs of feature vectors, whose (n, m) -th element is given by

$$W_{\mathcal{R}}^{nm} = \exp \left\{ -\frac{d_{\mathbf{C}}^2(\mathbf{h}_r^{(i)}, \mathbf{h}_\rho^{(j)})}{\varepsilon} \right\} \quad (8)$$

where ε is the kernel scale set according to [20] and $n = rL + i$ and $m = \rho L + j$. Let \mathbf{D} be a diagonal normalization matrix whose diagonal elements are $D^{nn} = \sum_m W_{\mathcal{R}}^{nm}$. Let $\tilde{\mathbf{W}}_{\mathcal{R}} = \mathbf{D}^{-1/2} \mathbf{W}_{\mathcal{R}} \mathbf{D}^{-1/2}$ be a normalized kernel that shares the eigenvectors with the normalized graph-Laplacian $\mathbf{I} - \tilde{\mathbf{W}}_{\mathcal{R}}$ [21]. It can be shown that the eigenvectors $\tilde{\varphi}_k$ of $\tilde{\mathbf{W}}_{\mathcal{R}}$ reveal the underlying structure of the data [22–25]. In the following, we assume that the decay rate of the room is the most significant underlying parameter of the decay rates of the observed reverberant speech signal. As a result, we obtain that the principal eigenvector¹ of length $L|\mathcal{R}|$ represents

¹The nontrivial eigenvector corresponding to the largest eigenvalue.

the L perturbed decay rates of the $|\mathcal{R}|$ training rooms, where $|\mathcal{R}|$ denotes the cardinality of \mathcal{R} . In particular, the n -th coordinate of the principal eigenvector relates to the decay rate as

$$\tilde{\varphi}_1^n = \phi(\lambda_r^{(i)}) \quad (9)$$

where $n = rL + i$ and ϕ is a monotonic function. Thus, the principal eigenvector organizes the feature vectors according to the values of the decay rates of the rooms up to a monotonic scaling. Furthermore, since the decay rates of the training rooms are known, we are able to use them for calibrating the values of the eigenvectors to the values of the decay rates, as described in Sec. 5.

5. ESTIMATING THE REVERBERATION TIME

Let \mathcal{U} denote a collection of “unseen” rooms with unknown RTs. From each such unseen room $u \in \mathcal{U}$, we obtain a *single* reverberant speech recording $x_u(n)$. Based on the reverberant speech, we compute the histograms (feature vectors) \mathbf{h}_u for $u \in \mathcal{U}$ of the decay rates of the energy envelopes of the signal in short time frames. In this section, we present the simultaneous estimation of the RTs of all the unseen rooms, which includes the case of a single unseen room as well.

We define a non-symmetric distance function between feature vectors from the unseen rooms and the training rooms as

$$a_C^2(\mathbf{h}_u, \mathbf{h}_r^{(i)}) = (\mathbf{h}_u - \mathbf{h}_r^{(i)})^T \mathbf{C}_r^{-1} (\mathbf{h}_u - \mathbf{h}_r^{(i)}) \quad (10)$$

for each $r \in \mathcal{R}$, $u \in \mathcal{U}$, and $i = 1, \dots, L$. Similarly to (8), we define a corresponding non-symmetric affinity matrix \mathbf{A} using a Gaussian as

$$A^{un} = \exp \left\{ -\frac{a_C^2(\mathbf{h}_u, \mathbf{h}_r^{(i)})}{\varepsilon} \right\} \quad (11)$$

where $n = rL + i$. We note that the construction of \mathbf{A} relies merely on the observed and training data, and does not use the unavailable covariance matrix of the feature vectors from the unseen rooms. Let $\tilde{\mathbf{A}} = \mathbf{D}_A^{-1} \mathbf{A} \boldsymbol{\omega}^{-1}$, where \mathbf{D}_A is a diagonal matrix whose diagonal elements are the sums of rows of \mathbf{A} , and $\boldsymbol{\omega}$ is a diagonal matrix whose diagonal elements are the sums of columns of $\mathbf{D}_A^{-1} \mathbf{A}$. In [23, 26], it is shown that

$$\mathbf{W}_{\mathcal{R}} = \tilde{\mathbf{A}}^T \tilde{\mathbf{A}}. \quad (12)$$

Thus, the eigenvectors φ_k of $\mathbf{W}_{\mathcal{R}}$, which represent the training rooms, can be obtained from the right singular vectors of $\tilde{\mathbf{A}}$.

Define a new affinity matrix between feature vectors of the unseen rooms as $\mathbf{W}_{\mathcal{U}} = \tilde{\mathbf{A}} \tilde{\mathbf{A}}^T$. The principal eigenvector of $\mathbf{W}_{\mathcal{U}}$ represents the underlying desired decay rates of the unseen rooms. In addition, the relationship between the eigenvectors of $\mathbf{W}_{\mathcal{U}}$ and $\mathbf{W}_{\mathcal{R}}$ is conveyed by the singular vector decomposition (SVD) of $\tilde{\mathbf{A}}$. By definition of the SVD, we obtain

$$\psi_k = \frac{1}{\sqrt{\mu_k}} \tilde{\mathbf{A}} \varphi_k \quad (13)$$

where μ_k is the k -th eigenvalue of $\mathbf{W}_{\mathcal{U}}$ and ψ_k is the k -th eigenvector of $\mathbf{W}_{\mathcal{U}}$ (and the left singular vector of $\tilde{\mathbf{A}}$). Thus, for $k = 1$, we obtain the extension of the representation of the underlying desired decay rates of the unseen rooms.

The above analysis specifies an efficient algorithm, presented in Algorithm 1, to obtain the representation of the underlying decay

Algorithm 1 Representation of the Decay Rates

Training Stage:

1. Obtain several recordings of reverberant speech from various training rooms.
2. Compute the decay rates of the reverberant speech signals in short time frames.
3. For each recording, compute a histogram of the obtained decay rates.
4. For each training room, compute the covariance matrix of the histograms of different recordings from this room.
5. Build a kernel between the histograms using (6) and (8).
6. Apply EVD on the kernel and view the values of its principal eigenvector as data-driven representation of the underlying decay rates of the training rooms.

Testing Stage for a Single Room:

1. Obtain a single recording of reverberant speech signals from an unseen room.
 2. Compute the decay rates in short time frames and the corresponding histogram.
 3. Build the non-symmetric kernel between the newly acquired histogram and the training histograms using (10) and (11).
 4. Extend the representation according to (13) to obtain a representation of the decay rate of the unseen room.
-

rates of unseen rooms. We note that the testing stage can be computed efficiently, circumventing the extensive computation of EVD and SVD of large kernel matrices [27].

We are able to obtain a natural extension of (9), thereby, recovering the decay rates of the unseen rooms up to a monotonic function. This means that we are able to organize the unseen rooms according to the values of their decay rates based solely on reverberant speech recordings from these rooms without the knowledge of the true decay rates (RTs). Exploiting this property to the full extent is beyond the scope of this paper and will be examined in future publications.

In this work, the decay rates (RTs) of all the training rooms are known and can be used to estimate the decay rates (RTs) of the unseen rooms. The SVD definition in (13) expresses the relationship between the representation of the decay rates of the training and unseen rooms. Since the true decay rates of the training rooms are known, we exploit the same relationship to estimate the decay rates of the unseen rooms. Substituting the training decay rates into (13) and setting $k = 1$ yields

$$\lambda_{\mathcal{U}} = \frac{1}{\sqrt{\mu_1}} \tilde{\mathbf{A}} \lambda_{\mathcal{R}} \quad (14)$$

where $\lambda_{\mathcal{R}}$ and $\lambda_{\mathcal{U}}$ are vectors consisting of the known decay rates of the training rooms and the obtained estimates of the decay rates of the unseen rooms, respectively.

It is worthwhile noting that the non-symmetric kernel implies an implicit probabilistic model in the features domain [18, 27]. We can redefine $\tilde{\mathbf{A}}$ as

$$\tilde{A}^{un} = \Pr(\mathbf{h}_u | \mathbf{h}_u \in \mathcal{H}_r) \quad (15)$$

where \mathcal{H}_r represents the local statistical model induced by the training pair $(\bar{\mathbf{h}}_r, \mathbf{C}_r)$. In other words, the (u, n) -th element of the non-symmetric kernel measures the probability to observe the feature vector \mathbf{h}_u given that it is associated with the “local” model induced

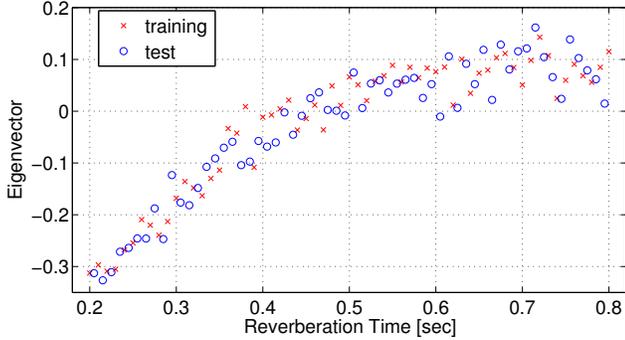


Fig. 1. A scatter plot of the obtained embedding in the noiseless case. Each cross/circle in the figure represents a single recording of reverberant speech. The y-axis depicts the values of the principal eigenvector and the x-axis depicts the true RTs.

by $(\bar{\mathbf{h}}_r, \mathbf{C}_r)$. Furthermore, (10) and (11) imply that this probability is normal with $\bar{\mathbf{h}}_r$ mean and \mathbf{C}_r covariance. Thus, the training set defines a multi-Gaussian mixture model in the features domain, where each Gaussian in the mixture is associated with a training room.

6. EXPERIMENTAL RESULTS

The experimental setup included 40 simulated training rooms with RTs evenly distributed between 0.2 and 0.8 s. For each training room, 30 training recordings of speech sampled at 16 kHz from arbitrary locations were used. Each recording consisted of a different speech signal (several speakers were used, including both females and males) of 4 s duration. In each recording, the speech signal was convolved with the RIR, simulated according to the image method [28] using [29], to obtain the reverberant speech. In the test stage, an “unseen” room with random RT ranging between 0.2 and 0.8 s was simulated and a recording of a speech signal (different from the signals used for training) was used. The decay rates of the reverberant speech were estimated in the time-frequency domain according to [10]. Short time frames of 256 samples with 75% overlap were used. The estimates were then averaged over the frequency bins to yield a single decay rate estimate for each time frame. In all the tested cases, the source-microphone distance was 2 m to attain a direct-to-reverberant energy ratio smaller than 0 dB.

Figure 1 presents a scatter plot of the obtained embedding. Each cross/circle in the figure represents a single recording of reverberant speech. We display 40 training recordings (one from each training room) and 100 test recordings marked by a cross and a circle, respectively. The y-axis depicts the values of the principal eigenvector ψ_1 in the corresponding coordinates and the x-axis depicts the true RTs. We observe that the principal eigenvector attaches a value for each reverberant speech recording, which is related to the true RT via a monotonic function. In particular, we observe that this function is approximately linear for RTs under 0.5 s.

In Fig. 2, we present a scatter plot of the estimated RTs of 100 test recordings as a function of the true RTs. We observe accurate estimations with a small variance when the RT is shorter than 0.4 s and a larger estimation variance for longer RTs. This result coincides with the obtained embedding depicted in Fig. 1.

To test noise robustness, we repeat the experiment 5 times for all rooms with RTs ranging from 0.2 until 0.8 s with an additive white Gaussian noise corrupting the measured signal under signal-

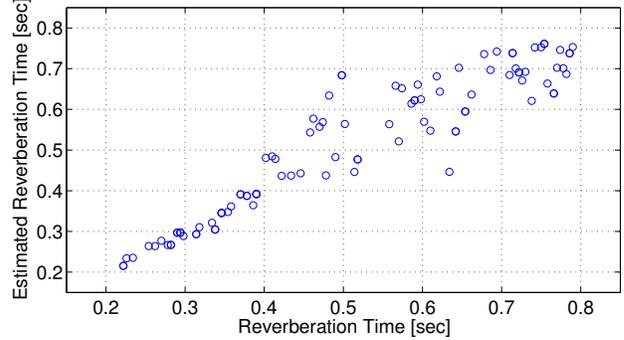


Fig. 2. A scatter plot of the estimated RT of test recordings as a function of the true RT in the noiseless case.

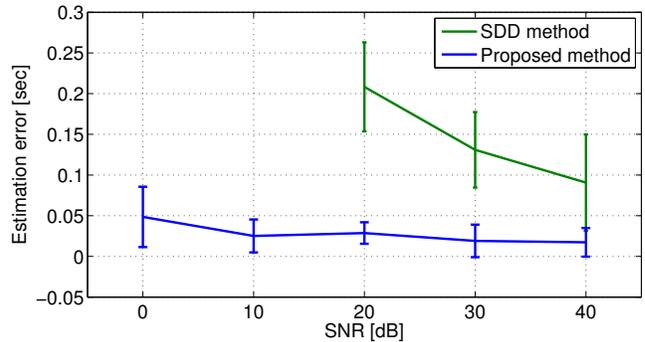


Fig. 3. Error bars summarizing the mean squared estimation error and the corresponding standard deviation.

to-noise ratio (SNR) conditions of 0, 10, 20, 30 and 40 dB, respectively. We compare the proposed method to the spectral decay distribution (SDD) method proposed in [10]. The same experimental setup and the same training and test signals are used in both methods. Figure 3 shows error bars representing the mean squared estimation error and the corresponding standard deviation obtained under the different SNR conditions. The obtained results of the competing method under high SNR conditions are comparable to the results reported in [10] for the noiseless case. We observe that the proposed method exhibits lower estimation error and variance than the competing method. In addition, the proposed method exhibits small degradation in performance as the SNR decreases whereas the competing method is more sensitive to measurement noise. The results obtained for the competing method with an SNR smaller than 20 dB were found incomparable and are therefore omitted.

7. CONCLUSIONS

A novel method for blindly estimating the RT based on recordings of noisy reverberant speech signals was proposed. An intrinsic geometric representation of the recordings is recovered using a kernel. This kernel consists of an affinity metric between the entire distributions of the decay rates of the recordings. We show that the obtained geometric representation can be used to accurately estimate the underlying RTs. In contrast to previous methods that were found sensitive to additive noise, the proposed method is data-driven, i.e., restrains from any model assumptions, and is shown to be much more robust to additive noise.

8. REFERENCES

- [1] W. C. Sabine, *Collected Papers on acoustics (Originally 1921)*, Peninsula Publishing, 1993.
- [2] H. Kuttruff, *Room Acoustics*, Taylor & Francis, London, fourth edition, 2000.
- [3] L. Couvreur and C. Couvreur, “On the use of artificial reverberation for ASR in highly reverberant environments,” in *Proc. 2nd IEEE Benelux Signal Processing Symposium (SPS-2000)*, Hilvarenbeek, The Netherlands, Mar. 2000, IEEE, pp. S001–S004.
- [4] R. Maas, E.A.P. Habets, A. Sehr, and W. Kellermann, “On the application of reverberation suppression to robust speech recognition,” in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 297–300.
- [5] E. A. P. Habets, *Single- and Multi-Microphone Speech Dereverberation using Spectral Enhancement*, Ph.D. thesis, Technische Universiteit Eindhoven, 2007.
- [6] E. A. P. Habets, S. Gannot, and I. Cohen, “Late reverberant spectral variance estimation based on a statistical mode,” *IEEE Signal Process. Lett.*, vol. 16, no. 9, pp. 770–774, Sept. 2009.
- [7] M. R. Schroeder, “Integrated-impulse method measuring sound decay without using impulses,” *J. Acoust. Soc. Am.*, vol. 66, no. 2, pp. 497–500, 1979.
- [8] T. J. Cox, F. Li, and P. Darlington, “Extracting room reverberation time from speech using artificial neural networks,” *Journal Audio Eng. Soc.*, vol. 49, no. 4, pp. 219–230, 2001.
- [9] R. Ratnam, D. L. Jones, and W. D. O’Brien, Jr., “Fast algorithms for blind estimation of reverberation time,” *IEEE Signal Process. Lett.*, vol. 11, no. 6, pp. 537–540, June 2004.
- [10] J. Y. C. Wen, E. A. P. Habets, and P. A. Naylor, “Blind estimation of reverberation time based on the distribution of signal decay rates,” in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Las Vegas, USA, Apr. 2008.
- [11] H. W. Löllmann and P. Vary, “Estimation of the reverberation time in noisy environments,” in *Proc. Intl. Workshop Acoust. Echo Noise Control (IWAENC)*, Sept. 2008, pp. 1–4.
- [12] H. W. Löllmann, E. Yilmaz, M. Jeub, and P. Vary, “An improved algorithm for blind reverberation time estimation,” in *Proc. Intl. Workshop Acoust. Echo Noise Control (IWAENC)*, Tel-Aviv, Israel, Aug. 2010.
- [13] H. W. Löllmann and P. Vary, “Estimation of the frequency dependent reverberation time by means of warped filter-banks,” in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 309–312.
- [14] N. López, Y. Grenier, G. Richard, and I. Bourmeyster, “Low variance blind estimation of the reverberation time,” in *Proc. Intl. Workshop Acoust. Signal Enhancement (IWAENC)*, Aachen, Germany, Sept. 2012.
- [15] N. D. Gaubitch, H. W. Löllmann, M. Jeub, T. H. Falk, P. A. Naylor, P. Vary, and M. Brookes, “Performance comparison of algorithms for blind reverberation time estimation from speech,” in *Proc. Intl. Workshop Acoust. Signal Enhancement (IWAENC)*, Aachen, Germany, Sept. 2012.
- [16] J. D. Polack, *La transmission de l’énergie sonore dans les salles*, Ph.D. thesis, Université du Maine, Le Mans, France, 1988.
- [17] R. Talmon and R. R. Coifman, “Differential stochastic sensing: Intrinsic modeling of random time series with applications to nonlinear tracking,” *submitted, Tech. Rep. YALEU/DCS/TR-1451*, 2012.
- [18] R. Talmon and R. R. Coifman, “Empirical intrinsic modeling of signals and information geometry,” *submitted, Tech. Rep. YALEU/DCS/TR-1467*, 2012.
- [19] R. Coifman and S. Lafon, “Diffusion maps,” *Appl. Comput. Harm. Anal.*, vol. 21, pp. 5–30, Jul. 2006.
- [20] M. Hein and J. Y. Audibert, “Intrinsic dimensionality estimation of submanifold in r^d ,” *ICML*, pp. 289–296, 2005.
- [21] F. R. K. Chung, *Spectral Graph Theory*, American Mathematical Society, 1997.
- [22] A. Singer and R. R. Coifman, “Non-linear independent component analysis with diffusion maps,” *Appl. Comput. Harm. Anal.*, vol. 25, no. 2, pp. 226–239, 2008.
- [23] 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.
- [24] R. Talmon, D. Kushnir, R. 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.
- [25] R. Talmon, I. Cohen, and S. Gannot, “Supervised source localization using diffusion kernels,” *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2011.
- [26] A. Haddad, D. Kushnir, and R. R. Coifman, “Filtering via a reference set,” *Tech. Rep. YALEU/DCS/TR-1441*, Feb. 2011.
- [27] 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, Nov. 2012.
- [28] 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, Apr. 1979.
- [29] E. A. P. Habets, “Room impulse response generator,” *Technische Universiteit Eindhoven, Tech. Rep.*, 2006.