Late Reverberant Spectral Variance Estimation Based on a Statistical Model

Emanuël A. P. Habets, *Member, IEEE*, Sharon Gannot, *Senior Member, IEEE*, and Israel Cohen, *Senior Member, IEEE*

***Abstract*—In speech communication systems the received microphone signals are degraded by room reverberation and ambient noise that decrease the fidelity and intelligibility of the desired speaker. Reverberant speech can be separated into two components, viz. early speech and late reverberant speech. Recently, various algorithms have been developed to suppress late reverberant speech. One of the main challenges is to develop an estimator for the so-called late reverberant spectral variance (LRSV) which is required by most of these algorithms. In this letter a statistical reverberation model is proposed that takes the energy contribution of the direct-path into account. This model is then used to derive a more general LRSV estimator, which in a particular case reduces to an existing LRSV estimator. Experimental results show that the developed estimator is advantageous in case the source-microphone distance is smaller than the critical distance.**

***Index Terms*—Reverberation suppression, speech enhancement, statistical room acoustics.**

I. Introduction

ACOUSTIC signals radiated within a room are linearly distorted by reflections from walls and other objects. This type of distortion is commonly known as reverberation and degrades the fidelity and intelligibility of speech, and the recognition performance of automatic speech recognition systems. In general, the degradation increases when the distance between the source and the microphone increases. One important effect of reverberation on speech is the lengthening of speech phonemes, which is caused by late reflections. Consequently, reverberation of one phoneme overlaps subsequent phonemes. Evidence has been found that this phenomenon, which is referred to as *overlap-masking*, decreases speech intelligibility [1].

Most reverberation suppression methods are designed to suppress late reverberation or in other words to estimate the early speech component and assume that the early and late reverberant speech component are mutually independent. The suppression is commonly carried out in the short-time Fourier transform (STFT) domain using so-called spectral enhancement methods. In order to perform spectral enhancement an estimate of the short-term power spectral density (or in the context of statistical spectral enhancement methods, spectral variance) of the late reverberant speech component is required. Hence, the main challenge is to estimate the spectral variance of the late reverberant speech component from the reverberant microphone signal. In the last decade several late reverberant spectral variance (LRSV) estimators have been developed [2]–[4]. Most LRSV estimators are based on statistical reverberation models that are formulated in the time domain.

The LRSV estimator developed in [2] is derived under the implicit assumption that the source-microphone distance is larger than the critical distance (i.e., the distance at which the direct-path energy and the reverberant energy are equal). When the source-microphone distance is smaller than the critical distance this LRSV estimator significantly overestimates the LRSV, as will be shown later. In this contribution we propose a statistical reverberation model in the STFT domain that takes into account the energy contribution of the direct-path and reverberation. A degenerated version of this model was recently presented in [5]. Subsequently, the proposed model is used to derive a novel LRSV estimator.

The letter is organized as follows: In Section II the problem is formulated. In Section III we propose a statistical reverberation model in the STFT domain. This model is used in Section IV to derive an estimator for the LRSV. The performance of the developed estimator is evaluated in Section V. Finally, conclusions are provided in Section VI.

II. Problem Formulation

The reverberant signal results from the convolution of the anechoic speech signal $s(n)$ and a causal AIR $h(n)$. Here we assume that the AIR is time-invariant and that its length is infinite. The reverberant speech signal at discrete-time n can be written as

$$z(n) = \sum_{n'=0}^{\infty} h(n')\, s(n-n'). \tag{1}$$

In the STFT domain the signal $s(n)$ is given by

$$S(\ell,k) = \sum_{n=-\infty}^{\infty} s(n)\, \tilde{\psi}(n-\ell R)\, \mathrm{e}^{-j2\pi/Nk(n-\ell R)} \tag{2}$$

where ℓ is the frame index, k is the frequency band index, R is the discrete time shift, and $\tilde{\psi}(m)$ denotes the analysis window of length N. Subsequently we can express $z(n)$ in the STFT domain as [6]

$$Z(\ell,k) = \sum_{k'=0}^{N-1} \sum_{\ell'=-\infty}^{\infty} H(\ell',k,k')\, S(\ell-\ell',k') \tag{3}$$

Manuscript received January 18, 2009; revised April 09, 2009. First published June 10, 2009; current version published July 09, 2009. This work was supported by the Israel Science Foundation (Grant 1085/05). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Kainam Thomas Wong.

E. A. P. Habets is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa, Israel, and also with the School of Engineering, Bar-Ilan University, Ramat-Gan, Israel (e-mail: e.habets@ieee.org).

S. Gannot is with the School of Engineering, Bar-Ilan University, Ramat-Gan, Israel (e-mail: gannot@eng.biu.ac.il).

I. Cohen is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa, Israel (e-mail: icohen@ee.technion.ac.il).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2009.2024791

where k and k' denote the band and cross-band frequency bin indices, respectively. The STFT response $H(\ell', k, k')$ is related to the impulse response $h(n)$ by

$$H(\ell', k, k') = [h(n) * \vartheta(n, k, k')]|_{n=\ell' R} \tag{4}$$

where $*$ denotes convolution with respect to n. The function $\vartheta(n, k, k')$ is related to the analysis window $\tilde{\psi}(m)$ and the synthesis window $\psi(m)$ of length N by [6]

$$\vartheta(n, k, k') = \mathrm{e}^{j2\pi/N k' n} \sum_{n'=-\infty}^{\infty} \tilde{\psi}(n') \psi(n' + n) \mathrm{e}^{-j2\pi/N n'(k-k')}. \tag{5}$$

To simplify the following discussion, and without loss of generality, it is assumed that the direct sound arrives at time instance n. Since our objective is to suppress late reverberation we split the AIR into two components such that

$$H(\ell, k, k') = \begin{cases} H_e(\ell, k, k'), & \text{for } 0 \leq \ell < N_e; \\ H_\ell(\ell, k, k') & \text{for } N_e \leq \ell \leq \infty; \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where $H_e(\ell, k, k')$ models the direct path and a few early reflections and $H_\ell(\ell, k, k')$ models all later reflections. The parameter $N_e (N_e \geq 1)$ defines which portion of the AIR is considered as late reverberation. Specifically, it is the time (measured with respect to the arrival time of the direct sound) at which we assume that the late reverberation starts.

Using (6) we can write the reverberant signal $Z(\ell, k)$ as

$$Z(\ell, k) = Z_e(\ell, k) + Z_\ell(\ell, k) \tag{7}$$

where

$$Z_e(\ell, k) = \sum_{k'=0}^{N-1} \sum_{\ell'=-\infty}^{N_e-1} H(\ell', k, k')\, S(\ell - \ell', k') \tag{8}$$

denotes the early spectral speech component and

$$Z_\ell(\ell, k) = \sum_{k'=0}^{N-1} \sum_{\ell'=N_e}^{\infty} H(\ell', k, k')\, S(\ell - \ell', k') \tag{9}$$

denotes the late reverberant spectral speech component.

An estimate of $Z_e(\ell, k)$ can be obtained using spectral enhancement methods given the LRSV $\lambda_\ell(\ell, k) = \mathrm{E}\left\{|Z_\ell(\ell, k)|^2\right\}$, where $\mathrm{E}\{\cdot\}$ denotes the mathematical expectation. Ideally, we would require $H(\ell, k, k')$ to estimate $\lambda_\ell(\ell, k)$. In practice $H(\ell, k, k')$ is not *a priori* known and blindly estimating $H(\ell, k, k')$ remains a difficult task. In order to avoid the need of estimating $H(\ell, k, k')$ our objective is to derive a direct estimator for $\lambda_\ell(\ell, k)$. In Section III we propose a statistical model for $H(\ell, k, k')$ that depends on a small parameter set. In Section IV an estimator for $\lambda_\ell(\ell, k)$ is derived using this statistical model.

III. Statistical Reverberation Model

Two commonly used statistical reverberation models are proposed by Schroeder [7] and Polack [8]. Schroeder derived several statistical properties of the acoustic transfer function, while Polack described the AIR as one-realization of a zero-mean Gaussian random sequence multiplied by an exponentially decaying function. In [9] Jot *et al.* showed that both models can be combined in the time-frequency domain. The combined model is valid after a certain mixing-time and above the Schroeder frequency, which is defined as the crossover frequency that marks the transition from well-separated to overlapping normal modes. It is important to notice that both models only describe the reverberant part of the response. Polack's model was used in [2] to derive an LRSV estimator. Consequently, the estimator does not take the energy contribution of the direct sound into account. In case the direct sound becomes dominant the LRSV will be overestimated, as will be shown later. Here we propose a statistical model for the band-to-band filters ($k' = k$) in the STFT domain that does take the direct path into account[1]:

$$H(\ell, k) \triangleq H(\ell, k, k) = \begin{cases} B_\mathrm{d}(k), & \text{for } \ell = 0; \\ B_r(\ell, k)\, \mathrm{e}^{-\alpha(k)\ell R}, & \text{for } \ell \geq 1 \end{cases} \tag{10}$$

where $\alpha(k)$ denotes the decay rate that is related to the reverberation time, and $B_\mathrm{d}(k)$ and $B_r(\ell, k)$ are zero-mean mutually independent and identically distributed (i.i.d.) Gaussian random variables. Accordingly, we have

1) $\mathrm{E}\{H(\ell_1, k_1) H^*(\ell_2, k_2\} = 0$ for $\ell_1 \neq \ell_2$ and $\forall k_1,\ k_2$;
2) $\mathrm{E}\{H(\ell, k_1) H^*(\ell, k_2)\} = 0$ for $k_1 \neq k_2$ and $\forall \ell$,

where $(\cdot)^*$ denotes complex conjugation. It is extremely interesting to note that different realization of $H(\ell, k)$ can be interpreted as different observation at different source and/or microphone positions in the enclosure.

Let us define $\beta_\mathrm{d}(k) = \mathrm{E}\{|B_\mathrm{d}(k)|^2\}$ and $\beta_r(k) = \mathrm{E}\{|B_r(\ell, k)|^2\}$. Now we can calculate the spectral variance (also known as spectral envelope) in the STFT domain

$$\begin{aligned}\lambda_h(\ell, k) &= \mathrm{E}\{|H(\ell, k)|^2\} \\ &= \begin{cases} \beta_\mathrm{d}(k), & \text{for } \ell = 0; \\ \beta_r(k)\, \mathrm{e}^{-2\alpha(k)\ell R}, & \text{for } \ell \geq 1 \end{cases}\end{aligned} \tag{11}$$

where $\alpha(k)$ is linked to the frequency dependent reverberation time $T_{60}(k)$ through

$$\alpha(k) \triangleq \frac{3\, \log_e(10)}{T_{60}(k)\, f_s} \tag{12}$$

where f_s denotes the sampling frequency in Hz.

According to the proposed model the energy of the direct-path is given by $\beta_\mathrm{d}(k)$ and the total energy of all reflections is given by

$$\sum_{\ell=1}^{\infty} \mathrm{E}\left\{|H(\ell, k)|^2\right\} = \beta_r(k)\, \frac{\mathrm{e}^{-2\alpha(k)R}}{1 - \mathrm{e}^{-2\alpha(k)R}}. \tag{13}$$

Therefore, we can express the direct-to-reverberation ratio (DRR) of $H(\ell, k)$ in dB as

$$\mathrm{DRR}(k) = 10 \log_{10} \left(\frac{1 - \mathrm{e}^{-2\alpha(k)R}}{\mathrm{e}^{-2\alpha(k)R}} \frac{\beta_\mathrm{d}(k)}{\beta_r(k)} \right). \tag{14}$$

IV. Late Reverberant Spectral Variance Estimator

In [5], a LRSV variance estimator was derived using a degenerated version of the statistical model and the assumption that $\kappa(k) = \beta_r / \beta_\mathrm{d} = 1$. Here we derive an estimator for the LRSV $\lambda_\ell(\ell, k)$ for $0 < \kappa(k) \leq 1$. In the following we assume that the spectral coefficients of the speech signal can be modelled as zero-mean i.i.d. complex random variables with a certain distribution and variance $\lambda_s(\ell, k)$. Using (3), the statistical

[1]This model is closely related to the time-domain model proposed in [3].

reverberation model described by (10) and the fact that $H(\ell,k)$ and $S(\ell,k)$ are mutually independent, we can express the reverberant spectral variance $\lambda_z(\ell,k) = \mathrm{E}\left\{|Z(\ell,k)|^2\right\}$ as

$$\begin{aligned}\lambda_z(\ell,k) &= \mathrm{E}\left\{\left|\sum_{\ell'=0}^{\infty} H(\ell',k)\, S(\ell-\ell',k')\right|^2\right\}\\ &= \sum_{\ell'=0}^{\infty} \lambda_h(\ell',k)\,\lambda_s(\ell-\ell',k). \qquad (15)\end{aligned}$$

Using (11) we can write (15) as

$$\lambda_z(\ell,k) = \lambda_{\mathrm{d}}(\ell,k) + \lambda_r(\ell,k) \qquad (16)$$

with

$$\lambda_{\mathrm{d}}(\ell,k) = \beta_{\mathrm{d}}(k)\,\lambda_s(\ell,k) \qquad (17)$$

and

$$\begin{aligned}\lambda_r(\ell,k) &= \sum_{\ell'=1}^{\infty} \beta_r(k)\,\mathrm{e}^{-2\alpha(k)R\ell'}\,\lambda_s(\ell-\ell',k)\\ &= \mathrm{e}^{-2\alpha(k)R}\\ &\quad\times\left[\lambda_r(\ell-1,k) + \beta_r(k)\,\lambda_s(\ell-1,k)\right]. \qquad (18)\end{aligned}$$

Now, using (17), the last expression can be rewritten as

$$\lambda_r(\ell,k) = \mathrm{e}^{-2\alpha(k)R}\left[\lambda_r(\ell-1,k) + \kappa(k)\,\lambda_{\mathrm{d}}(\ell-1,k)\right]. \qquad (19)$$

Finally, using (16) and reorganizing the terms, we obtain

$$\begin{aligned}\lambda_r(\ell,k) = &\left[1-\kappa(k)\right]\,\mathrm{e}^{-2\alpha(k)R}\,\lambda_r(\ell-1,k)\\ &+\kappa(k)\,\mathrm{e}^{-2\alpha(k)R}\,\lambda_z(\ell-1,k). \qquad (20)\end{aligned}$$

Hence, we can express the spectral variance $\lambda_r(\ell,k)$ of the reverberant component using $\kappa(k)$, $\alpha(k)$ and the spectral variance of the reverberant signal $Z(\ell,k)$. As seen from (20) we require that $0 < \kappa(k) \leq 1$.

The late reverberant spectral variance $\lambda_\ell(\ell,k)$ is given by

$$\lambda_\ell(\ell,k) = \mathrm{e}^{-2\alpha(k)R(N_e-1)}\,\lambda_r(\ell-N_e+1,k). \qquad (21)$$

In the following an instantaneous estimate of the reverberant spectral variance $\lambda_z(\ell,k)$ was used, which is given by $|Z(\ell,k)|^2$.

For $\alpha(k)=\alpha$ and $\kappa(k)=1$ the LRSV is given by

$$\lambda_\ell(\ell,k) = \mathrm{e}^{-2\alpha R N_e}\,\lambda_z(\ell-N_e,k) \qquad (22)$$

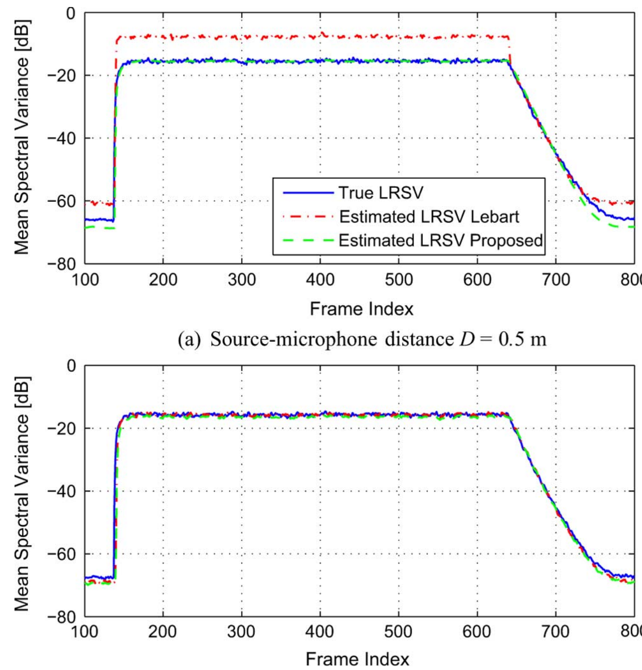
which is equivalent to the LRSV estimator proposed by Lebart *et al.* in [2].

V. Performance Evaluation

In this section, we evaluate the performance of the LRSV estimator. The STFT synthesis window is a Hamming window and the analysis window is the related bi-orthogonal window (cf. [6], [10]). The windows are of length $N = 128$ and the overlap between two successive STFT frames was 75% (i.e., $R = 32$). For $\kappa = 1$, N_e can be chosen equal or larger than the boundary point between early reflections and late reverberation. For $\kappa \ll 1$, a smaller value of N_e can be chosen depending on the subjective preference of the listener. Here the time instance was set to 48 ms, i.e., $N_e = 12$.The reverberation time $T_{60}(k)$ was determined for each octave band by applying Schroeder's method [11] to a bandpass filtered version of the AIR. The decay rate $\alpha(k)$ was calculated using (12). The parameter $\kappa(k)$ was calculated by solving (14) using

$$\widehat{\mathrm{DRR}}(k) = 10\log_{10}\left(\frac{\sum_{n=0}^{n_{\mathrm{d}}-1}[h(n)]^2}{\sum_{n=n_{\mathrm{d}}}^{\infty}[h(n)]^2}\right) \quad \forall k \qquad (23)$$

where $n_{\mathrm{d}} = 16$. In practice one can use blind estimation procedures as proposed in [2], [4], [12], [13].



(a) Source-microphone distance $D = 0.5$ m

(b) Source-microphone distance $D = 2.5$ m

Fig. 1. Mean spectral variance of i) the late reverberant signal $Z_\ell(\ell,k)$ (solid), ii) estimated using Lebart's estimator (dashed-dotted), iii) estimated using the proposed estimator (dashed).

A. Synthetic Signal

Two synthetic reverberant signals were created by convolving a white Gaussian noise pulse with two AIRs. The source-microphone distance $D = \{0.5, 2.5\}$ m, $T_{60} = 500$ ms and the room size is $6\times 8\times 5$ m (length $\times$ width $\times$ height). The AIRs were generated using an efficient implementation of the celebrated image method [14]. The LRSV was estimated using Lebart's estimator (or equivalently the proposed with $\kappa(k)=1$) and the proposed estimator. Here the 'true' LRSV $\lambda_\ell(\ell,k) = |Z_\ell(\ell,k)|^2$, where $Z_\ell(\ell,k)$ is calculated using (9).

The mean spectral variance was obtained by averaging over all N frequency bins. In Fig. 1, the resulting true and estimated LRSVs are depicted. For $D = 0.5$ m Lebart's estimator significantly overestimates the LSRV while the proposed estimator closely follows the true spectral variance. In addition, it should be noticed that scaling this estimate (as proposed in [4]) will

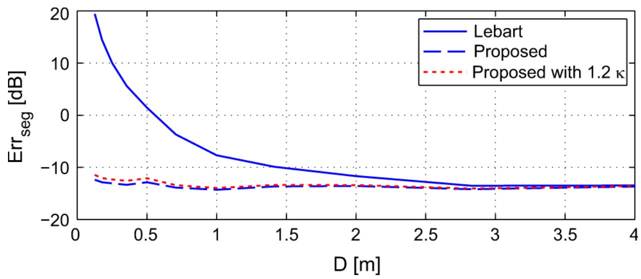


Fig. 2. Mean segmental error ($\mathrm{Err_{seg}}$) calculated using all speech signals from the APLAWD database for Lebart's (solid), the proposed estimator with $\hat{\kappa} = \kappa$ (dashed) and $\hat{\kappa} = 1.2\kappa$ (dotted) as a function of the source-microphone distance D.

correct the overestimation during the steady-state period while it will result in an underestimation of the LRSV during the free-decay (i.e., when the anechoic signal is not active).

B. Speech Signals

The performance of the LRSV estimator was tested using reverberant speech ($f_s = 8$ kHz) that was generated by convolving all anechoic speech fragments from the APLAWD database [15] with various AIRs. We analysed the segmental error of the LRSV estimated using Lebart's and the proposed estimator using

$$\mathrm{Err_{seg}} = \mathrm{mean}_\ell \left\{ \frac{\mathrm{mean}_k \left\{ \left[\lambda_\ell(\ell,k) - \hat{\lambda}_\ell(\ell,k) \right]^2 \right\}}{\mathrm{mean}_k \left\{ \left[\lambda_\ell(\ell,k) \right]^2 \right\}} \right\}. \quad (24)$$

In Fig. 2, the $\mathrm{Err_{seg}}$ is shown for $T_{60} = 500$ ms and source-microphone distances ranging between 0.125 and 4 m. To demonstrate the effect of estimation error of κ the $\mathrm{Err_{seg}}$ is also shown for $\hat{\kappa} = 1.2\,\kappa$. The analysis shows that the developed estimator is advantageous for all source-microphone distances.

C. Reverberation Suppression

As an example the log-spectral amplitude gain function [16] was used to suppress the late reverberant speech component. The gain function was lower-bounded to -12 dB. The *a priori* signal to reverberation ratio (SRR) required by the gain function was estimated using the decision-directed approach with a weighting factor of 0.98 (c.f [10]). To reduce the variance of the estimated LRSV a time-averaged version was used. The performance was evaluated using the segmental SRR and log spectral distance (LSD) measures (cf. [10]). For each reverberation time $T_{60} = \{0.35, 0.50, 0.65\}$ s the results were averaged over ten different source-microphone positions (with equal source-microphone distance $D = \{0.5, 3.5\}$ m), and 50 speech fragments. The mean segmental SRR and LSD results are summarized in Table I. For negative DDRs the performance in terms of the objective measures is almost equal, and the difference becomes smaller when the DRR decreases. For positive DRRs the proposed estimator gives an advantage over Lebart's estimator.

TABLE I
IMPROVEMENT OF THE SEGMENTAL SRR AND LSD FOR DIFFERENT REVERBERATION TIMES AND SOURCE-MICROPHONE DISTANCES

D [m]	T_{60} [s]	$\widehat{\mathrm{DRR}}$ [dB]	ΔsegSRR [dB] Lebart	ΔsegSRR [dB] Proposed	ΔLSD [dB] Lebart	ΔLSD [dB] Proposed
0.5	0.35	11.29	-2.18	0.85	-0.56	0.12
	0.50	8.95	-0.42	1.90	-0.22	0.39
	0.65	7.47	0.71	2.58	0.16	0.67
3.5	0.35	-3.62	3.30	3.11	1.36	1.32
	0.50	-6.48	4.51	4.45	2.39	2.37
	0.65	-8.22	5.12	5.12	3.12	3.14

VI. CONCLUSIONS

In this letter, a statistical reverberation model was proposed that explicitly takes the energy contribution of the direct-path and reverberation into account. Using this statistical model a novel LRSV estimator was derived that depends on the (frequency dependent) reverberation time and (frequency dependent) DRR. The performance evaluation demonstrated that the derived LRSV estimator is especially advantageous in case the source-microphone distance is smaller than the critical distance or in other words in case the DRR is positive.

REFERENCES

[1] A. K. Nábělek, T. R. Letowski, and F. M. Tucker, "Reverberant overlap- and self-masking in consonant identification," *J. Acoust. Soc. America*, vol. 86, no. 4, pp. 1259–1265, 1989.
[2] K. Lebart, J. M. Boucher, and P. Denbigh, "A new method based on spectral subtraction for speech dereverberation," *Acta Acoustica*, vol. 87, no. 3, pp. 359–366, 2001.
[3] E. A. P. Habets, "Single- and Multi-Microphone Speech Dereverberation Using Spectral Enhancement," Ph.D., Techn. Univ. Eindhoven, Eindhoven, The Netherlands, 2007.
[4] I. Tashev and D. Allred, "Reverberation reduction for improved speech recognition," in *Proc. Joint Workshop on Hands-Free Speech Communication and Microphone Arrays (HSCMA'05)*, Piscataway, NJ, Mar. 2005.
[5] E. A. P. Habets, S. Gannot, and I. Cohen, "Speech dereverberation using backward estimation of the late reverberant spectral variance," in *Proc. IEEE Conv. Electrical & Electronics Engineers in Israel (IEEEI)*, Israel, Dec. 2008, pp. 384–388.
[6] Y. Avargel and I. Cohen, "System identification in the short-time Fourier transform domain with crossband filtering," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 4, pp. 1305–1319, May 2007.
[7] M. R. Schroeder, "Frequency correlation functions of frequency responses in rooms," *J. Acoust. Soc. America*, vol. 34, no. 12, pp. 1819–1823, 1962.
[8] J. Polack, "La transmission de lenergie sonore dans les salles," Ph.D., Univ. Maine, Le Mans, France, 1988.
[9] J.-M. Jot, L. Cerveau, and O. Warusfel, "Analysis and synthesis of room reverberation based on a statistical time-frequency model," *Audio Eng. Soc., 103th Conv.*, Aug. 1997.
[10] I. Cohen and S. Gannot, "Spectral enhancement methods," in *Springer Handbook of Speech Processing*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds. New York: Springer-Verlag, 2007, ch. 45, pt. H.
[11] M. R. Schroeder, "New method of measuring reverberation time," *J. Acoust. Soc. America*, vol. 37, pp. 409–412, 1965.
[12] H. W. Löllmann and P. Vary, "Estimation of the reverberation time in noisy environments," in *Proc. Int. Workshop of Acoustic Echo and Noise Control (IWAENC)*, Sep. 2008, pp. 1–4.
[13] T. H. Falk and W.-Y. Chan, "Temporal dynamics for blind measurement of room acoustical parameters," *IEEE Trans. Instrum. Meas.*, to be published.
[14] E. A. P. Habets, Room Impulse Response (RIR) Generator May 2008 [Online]. Available: http://home.tiscali.nl/ehabets/rirgenerator.html
[15] G. Lindsey, A. Breen, and S. Nevard, SPAR's Archivable Actual-Word Databases, Univ. College London, London, U.K., 1987.
[16] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean square error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 2, pp. 443–445, Apr. 1985.