A blind algorithm for reverberation-time estimation using subband decomposition of speech signals

Thiago de M. Prego,[a)] Amaro A. de Lima,[b)] and Sergio L. Netto
Electrical Engineering Program, COPPE, Federal University of Rio de Janeiro, Avenue Athos da Silveira Ramos 149, Rio de Janeiro, RJ, 21941-972, Brazil

Bowon Lee, Amir Said, and Ronald W. Schafer
Hewlett-Packard Laboratories, 1501 Page Mill Road, Palo Alto, California 94304

Ton Kalker[c)]
Huawei Innovation Center US R&D, 2330 Central Expressway, Santa Clara, California 95050

(Received 7 June 2011; revised 26 January 2012; accepted 29 January 2012)

An algorithm for blind estimation of reverberation time (RT) in speech signals is proposed. Analysis is restricted to the free-decaying regions of the signal, where the reverberation effect dominates, yielding a more accurate RT estimate at a reduced computational cost. A spectral decomposition is performed on the reverberant signal and partial RT estimates are determined in all signal subbands, providing more data to the statistical-analysis stage of the algorithm, which yields the final RT estimate. Algorithm performance is assessed using two distinct speech databases, achieving 91% and 97% correlation with the RTs measured by a standard nonblind method, indicating that the proposed method blindly estimates the RT in a reliable and consistent manner.
 [http://dx.doi.org/10.1121/1.3688503]

PACS number(s): 43.55.Br, 43.60.Pt [NX] Pages: 2811–2816

I. INTRODUCTION

Reverberation is an acoustical effect occurring when several copies of a sound signal, with different delays and decreasing intensity levels, are perceived altogether. These copies are commonly due to signal reflections in an enclosure, which can vary in size, for instance, from our ear internal chamber (an important factor in hearing-aid devices[1]) to a large medieval cathedral.

Heavy amounts of reverberation can hinder speech intelligibility, possibly affecting the perceptual quality of a speech signal. The T_{60} reverberation time (RT) attempts to quantify the reverberation effect by specifying the time interval for a sound level to decay 60 dB after ceasing its stimulus.[2] A reliable RT estimation may be used to assess the acoustic characteristics of a room or to design a proper dereverberation scheme for a particular audio system.

The reverberation effect is often modeled by the convolution of the original anechoic source $s(n)$ with a length-N room impulse response (RIR) $h(n)$, generating the reverberating sound $s_r(n)$, as given by[3]

$$s_r(n) = \sum_{k=0}^{N-1} h(k)s(n-k). \tag{1}$$

This paper addresses the problem of estimating the T_{60} parameter from a single reverberant speech signal, $s_r(n)$, which is referred to as a blind or no-reference approach. Initial work on this particular subject includes Refs. 4 and 5, where the authors model the decaying process by an exponential function whose time constant is estimated using the entire reverberant signal. Later, Vieira[6] restricted the reverberation modeling process to the so-called free-decay regions (FDRs), which are the signal portions where the sound energy decreases consistently in several consecutive blocks. By doing so, one can achieve a better model fitting, thus improving the accuracy of the T_{60} estimate. A modified energy-decay model,[7] which also considers an additive noise component, was incorporated into the algorithm by Vieira in Ref. 8, making the RT estimate more robust to measurement noise.

Other work in blind RT estimation also includes Ref. 9, which uses a pitch-based RT model that restricts the analysis to a small T_{60} range; Ref. 10, which requires a quadratic mapping function highly dependent on the algorithm's implementation; and Ref. 11, which incorporates a noise-reduction stage to the algorithm described in Ref. 4, but still employs the entire signal, thus presenting a high-variance estimation process.

Although the FDR constraint improves upon the resulting RT estimate, it forces one to consider very long signals (more than 40 s, for instance, as in Refs. 6 and 8, alternating sound activity and pauses, to generate reliable statistics about the RT process. The proposed algorithm, which is also focused on the FDRs, mitigates the requirement of very long signals by performing a spectral decomposition on the reverberant signal, following the approach used in Refs. 12 and 13. The RT model can then be applied to each of the signal subbands, yielding a large number of partial RT estimates, even for a relatively short speech signal, making the final algorithm suitable for on-line applications.

[a)]Author to whom correspondence should be addressed. Electronic mail: thprego@lps.ufrj.br
[b)]Also at: Federal Center for Technological Education Celso Suckow da Fonseca (CEFET-RJ), Estrada de Adrianopolis 1317, Nova Iguaçu, RJ, 26041-271, Brazil.
[c)]Work was performed while at Hewlett-Packard Laboratories.

The proposed RT estimation algorithm is presented in Sec. II. Section III discusses some system design issues and evaluates the system performance using two distinct speech databases.

II. PROPOSED ALGORITHM

The proposed algorithm is comprised of four steps, which are detailed in Secs. II A–II D:

(1) Time-frequency representation of reverberant signal $s_r(n)$;
(2) Localization of FDRs in each subband;
(3) RT estimation for all detected subband FDRs;
(4) Statistical analysis of subband RT estimates to generate the final T_{60} estimate.

A. Time-frequency representation

In this initial stage, the reverberant speech signal, $s_r(n)$, is divided into frames using a length-M window function $w(n)$, and a discrete Fourier transform (DFT), $\mathcal{F}\{\cdot\}$, is applied to each frame, generating the time-frequency representation $S_r(k, l)$ such that

$$S_r(k,l) = \mathcal{F}\{w(n)s_r(n)\}, \tag{2}$$

for $k = 0, 1, \ldots, (K-1)$, $l = 0, 1, \ldots, (L-1)$, and $n = l(M-V), l(M-V)+1, \ldots, l(M-V)+M-1$, where K is the DFT length, L is the total number of speech frames, and V is the number of overlapping samples of two consecutive frames.

Since most of the speech energy lies within the analog frequency range $0 \leq f \leq 4$ kHz, we restrict all subsequent analyses to the values of k such that $0 \leq F_s k/K \leq 4$ kHz, thus achieving a more reliable RT estimate, where $F_s \geq 8$ kHz is the associated sampling frequency.

B. Subband FDR detection

As mentioned in Sec. I, the FDRs are characterized by a consistent energy drop in consecutive signal frames. In the proposed algorithm, however, this search must be performed for each individual subband, as these spectral components present a distinct energy pattern.[14] By defining the energy of the kth subband of the lth signal frame as

$$E(k,l) = |S_r(k,l)|^2, \tag{3}$$

the FDR search is performed across the frame index $l = 0,1,\ldots,(L-1)$, for each frequency bin k.

Extending Vieira's criterion[6,8] to the transform domain, a subband FDR may be characterized by a decrease in the value of $E(k, l)$ for a minimum of 500 ms along l within subband k. Using M samples/frame with V overlapping samples/frame, this 500-ms interval translates into consecutive

$$L_{\text{lim}} = \frac{0.500\,F_s}{M-V} \tag{4}$$

subband frames with decreasing energy. When using the values of $M = 0.05\,F_s$ and $V = M/4$, as determined in Sec. III B, leads to $L_{\text{lim}} \approx 13$. In the proposed algorithm, however, if no FDR satisfies this criterion in a given subband, this threshold number L_{lim} is reduced iteratively down to as low as 3 consecutive frame-energy decreases. This lower limit 3 for L_{lim} was determined empirically and guaranteed at least one FDR for each subband in all signals considered in this work; accepting less than 3 consecutive decays, however, would identify many false FDRs along a real speech signal. This small modification, of decreasing L_{lim} in case no FDR is found within a given subband, guarantees a minimum amount of meaningful data for the following stages of the algorithm.

The FDR detection process in a speech signal comprising two consecutive sentences is depicted in Fig. 1, where the horizontal dark lines in the upper plot indicate the resulting FDRs in each band. From this figure one can easily observe the distinct FDR pattern in each subband, with these

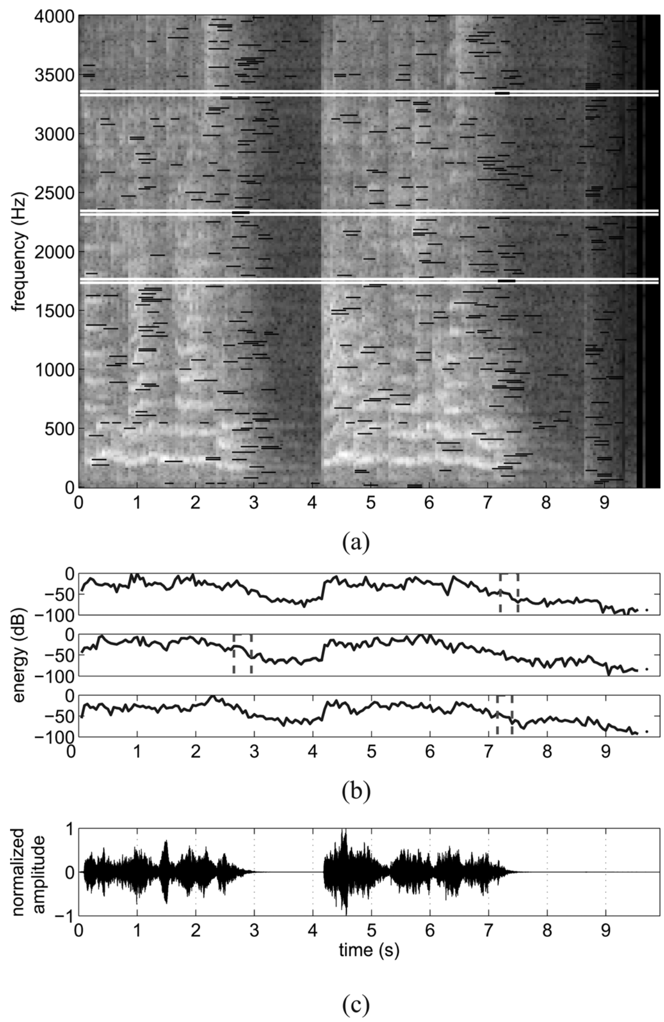

FIG. 1. Characterization of subband FDRs: (a) Spectrogram showing all subband FDRs (using $M = 0.05\,F_s$ and $V = M/4$) as dark thin lines; (b) three subband signals (identified by horizontal white lines in upper plot), with center frequencies at 1750, 2330, and 3340 Hz, respectively, showing corresponding FDRs within vertical dashed lines; (c) two-sentence speech signal.

FDRs concentrating in the beginning of the silence intervals, where the fullband reverberation process dominates.

C. Subband RT estimation

Standard algorithms estimate T_{60} as the time interval required by some linear fitting of the energy decay function (EDF)

$$c(n) = 10\log_{10}\left(\frac{\sum_{\nu=n}^{N-1} h^2(\nu)}{\sum_{\nu=0}^{N-1} h^2(\nu)}\right)\text{dB}, \tag{5}$$

for $n = 0,1,\ldots,(N-1)$, to drop 60 dB.[2,7,15] The key factor on most RT estimation algorithms is to find the time interval $n_1 \le n \le n_2$ that yields a reliable linear EDF approximation. The value of n_1 is commonly taken as the point where $c(n_1) = -5$ dB,[16] whereas n_2 is chosen in such a way that the resulting fitting yields the minimum mean-squared error (MSE). In general, the algorithms described in Refs. 7 and 15 tend to be very reliable in the presence of noise. However, these algorithms also demand a large number of EDF points to generate a reliable RT estimate, making them unpractical to our frame-based FDR processing.

Therefore, we employ here an extension of Schroeder's original algorithm[2] to subband signals, allowing one to base all subsequent processing on the subband-frame energy function $E(k,l)$ defined in Eq. (3). In this sense, the frame-based subband EDF (SEDF) is defined as

$$\bar{c}(k,l) = 10\log_{10}\left(\frac{\sum_{\lambda=n}^{\bar{L}-1} E(k,\lambda)}{\sum_{\lambda=0}^{\bar{L}-1} E(k,\lambda)}\right)\text{dB}, \tag{6}$$

for $l = 0,1,\ldots,(\bar{L}-1)$, where $\bar{L}$ is the number of frames within a subband FDR. The RT estimate is defined as the amount of time required by a linear fitting of the SEDF, performed within the interval $l_1 \le l \le l_2$, to drop 60 dB, with the extremes l_1 and l_2 chosen in a similar fashion as before.

When using real speech signals, one may not observe a consistent 60-dB decay in all SEDFs. In such cases, the linear fitting in Schroeder's algorithm considers only a reduced attenuation interval, corresponding to a range that is smaller than 60 dB, and the T_{60} RT value needs to be extrapolated. When dealing with frames instead of samples, the time resolution of l_1 and l_2 drops accordingly, increasing the variance of the RT estimate in a significant manner, particularly when l_2 is close to l_1. To minimize this effect, if a best linear fitting is such that $(\bar{c}(k,l_1) - \bar{c}(k,l_2)) < 10$ dB, we perform a new fitting using, whenever possible, l_2 such that $\bar{c}(k,l_2) = -65$, -45, -25, or -15 dB, in this particular order of preference. Starting at $\bar{c}(k,l_1) = -5$ dB, these noise-floor levels for $\bar{c}(k,l_2)$ lead to the values of T_{60}, T_{40}, T_{20}, and T_{10}, respectively, as defined in Ref. 16, which, by assuming a linear decay energy, can be readily converted into the desired RT scale.

D. Statistical analysis of subband RTs

Assuming that a total of R_k FDRs were found in the kth subband, each partial RT estimate can be denoted by $\hat{T}_{60}(r,k)$, for $r = 1,2,\ldots,R_k$. The final stage in the proposed algorithm is to sort out all these $\hat{T}_{60}(r,k)$ estimates to generate a final RT estimate $\hat{T}_{60}$.

Reference 4 employs several strategies to remove spurious partial estimates, which is not necessary in our case, since we restrict the analysis to the signal FDRs. In his algorithms,[6,8] Vieira defines $\hat{T}_{60}$ as the peak of a $\hat{T}_{60}(r,k)$ histogram, which, however, is highly dependent on the chosen histogram resolution.

In the proposed scheme, we first determine a subband estimate $\bar{T}_{60}$ as the median value of all subband medians $\bar{T}_{60}(k)$, thus avoiding biased/noisy extreme values. In fact, the median operator eliminates small (which do not affect the fullband dynamics significantly) and large (which may carry large estimation error) partial estimates, yielding a subband estimate that seems to represent the entire RT process in a reliable manner by presenting a large statistical correlation with the true RT value. However, when generating the $\bar{T}_{60}$ estimate, the median operator compresses the associated dynamic range, which must be compensated in the next stage of the algorithm to obtain the correct fullband RT.

The relationship between the subband ($\bar{T}_{60}$) and fullband ($\hat{T}_{60}$) RT estimates is quite difficult to model and constitutes an open problem in the associated literature.[10,13,17] Our subband RT estimates, for instance, although highly correlated to the standard T_{60} metric, vary within a different dynamic range due to the median operator employed in its derivation, thus requiring an additional mapping function, which in this work is described by

$$\hat{T}_{60} = \alpha\bar{T}_{60} + \beta, \tag{7}$$

with α and β chosen in a system training stage. For the values of $\alpha = 3.4$ and $\beta = -1170$ ms, as given in Sec. III C below, when the subband RT estimates vary, for instance, within the range $380 \le \bar{T}_{60} \le 640$, the associated fullband estimates will vary within $100 \le \hat{T}_{60} \le 1000$, representing a simple scale expansion of the RT dynamic range. It is important to stress that this mapping adjusts the subband measure $\bar{T}_{60}$ to the fullband signal RT without affecting the linear correlation with the theoretical RT process.

III. PERFORMANCE ASSESSMENT

A. Speech databases

Two databases of reverberant speech signals were employed to assess the performance of the proposed algorithm. The theoretical RT for each database was obtained using the non-blind algorithm described in Ref. 15.

(1) Database A: This database was developed using three different forms for imposing the reverberation effect:
   (a) Artificial reverberation: This method employed six artificially generated RIRs using the method of images, with RTs in the range of {200, 300, 400, 500, 600, 700} ms, emulating a source-microphone

TABLE I. Room characteristics for natural reverberation effect in Database A.

Room type	Dimensions [m × m × m]	$\tilde{T}_{60}$ [ms]	d [m]
Booth	3.0 × 1.8 × 2.2	120	0.5, 1, 1.5
Office	5.0 × 6.4 × 2.9	430	1, 2, 3
Meeting	8.0 × 5.0 × 3.1	230	1.45, 1.7, 1.9, 2.25, 2.8
Lecture	10.8 × 10.9 × 3.15	780	2.25, 4, 5.6, 7.1, 8.7, 10.2

distance $d_{SM} = 1.8$ m in a room of dimensions length × width × height = 4 × 3 × 3 m^3, as detailed in Ref. 18.

(b) Natural reverberation: This method employed RIRs obtained from the direct recordings in four distinct rooms with different RT characteristics and several source-microphone distances d for each room, as detailed in Table I.[19]

(c) Real reverberation: In this method, the degraded signals were directly recorded in seven distinct rooms, as summarized in Table II.

It must be made clear that "Natural reverberation" indicates convolution of measured RIRs (Ref. 19) and an anechoic signal, whereas "Real reverberation" refers to recording of signals in real rooms. Database A considered 4 anechoic speech signals (2 from a male speaker and 2 from a female speaker), resulting in 24 artificially degraded, 68 naturally degraded, and 108 signals degraded with the real reverberation approach, all sampled at $F_s = 48$ kHz.

(2) Database B: This corresponds to the MARDY database,[20] which includes 16 reverberant signals, recorded directly in an auditorium and their 16 dereverberated versions using the delay-and-sum algorithm, making a total of 32 speech signals with $F_s = 16$ kHz. The database considers 2 different speakers (1 male and 1 female), 4 values for the source-microphone distance $d = 1, 2, 3, 4$ m, and 2 types (reflective and absorbent) of wall panels, with RTs around 447 and 291 ms, respectively.

B. Algorithm adjustment

Database A was divided into two complementary databases, A_1 and A_2, of the same size and covering all reverberation effects present in the complete database. Database A_1 was then employed to perform some parameter adjustment in the proposed algorithm, whereas Databases A_2 and B were used to validate the overall algorithm performance.

TABLE II. Room characteristics for real reverberation effect in Database A.

Room type	Dimensions [m × m × m]	$\tilde{T}_{60}$ [ms]	d [m]
Booth	2.1 × 1.8 × 2.4	140	0.5, 1, 1.5
Office1	7.4 × 5.0 × 2.7	390	1, 2, 3, 4
Lecture1	15.0 × 10.0 × 4.0	570	1, 2, 3, 4
Meeting1	10.0 × 4.8 × 3.2	650	1, 2, 3, 4
Lecture2	16.5 × 8.2 × 3.5	700	1, 2, 3, 4
Meeting2	9.0 × 7.3 × 3.5	890	1, 2, 3, 4
Office2	7.4 × 4.8 × 4.3	920	1, 2, 3, 4

TABLE III. Statistical correlation between estimated and theoretical RTs for Database A with distinct values of frame size W for $\upsilon = 25\%$ of overlap percentage and $\bar{K} = 1024$-length DFT.

W [ms]	Database A_1	Database A_2
30	86.4	84.4
35	88.7	88.1
40	89.3	88.2
45	92.0	90.1
50	**92.1**	**91.0**
55	91.1	89.4
60	89.6	88.8
65	91.5	90.7
70	89.9	89.3
75	89.8	86.7
80	88.0	85.0
85	86.6	85.4
90	85.2	85.3
95	84.6	81.2
100	84.2	84.0

The parameters considered in this analysis are the frame duration ($W = M/F_s$), overlap percentage ($\upsilon = V/M \times 100\%$) in consecutive frames, and number $\bar{K}$ of DFT bins within the [(0, 4)] kHz band. Performance was assessed by the statistical correlation between estimated RTs using the proposed algorithm and the algorithm described in Ref. 15, as provided in Table III for $\upsilon = 25\%$ and $\bar{K} = 1024$ bins. Other values of $\upsilon = \{0, 50, 75\}$ and $\bar{K} = \{512, 2048\}$ were also considered in additional experiments, without any improvement in system performance. Based on the results summarized in Table III, the block length was chosen as $W = 50$ ms, which yielded a 92% correlation score for Database A_1.

C. Validation stage

The algorithm performance for Database A_2 is also shown in Table III, where one observes a 91% correlation score achieved by the adjusted algorithm with nontraining data.

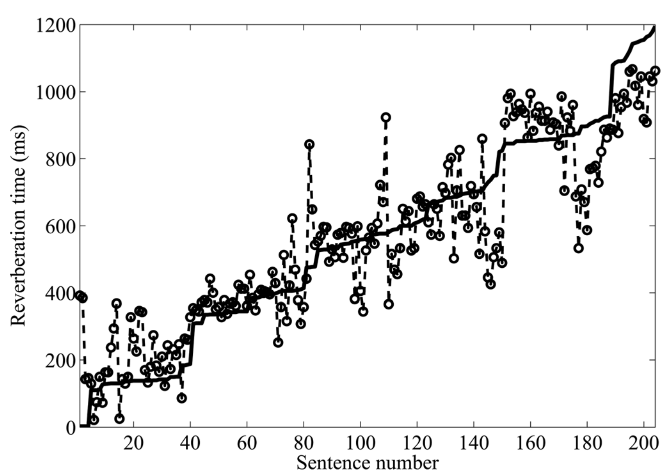

FIG. 2. Estimated RT values using proposed blind (dashed line) and reference non-blind (solid line) methods for all 204 signals in Database A.

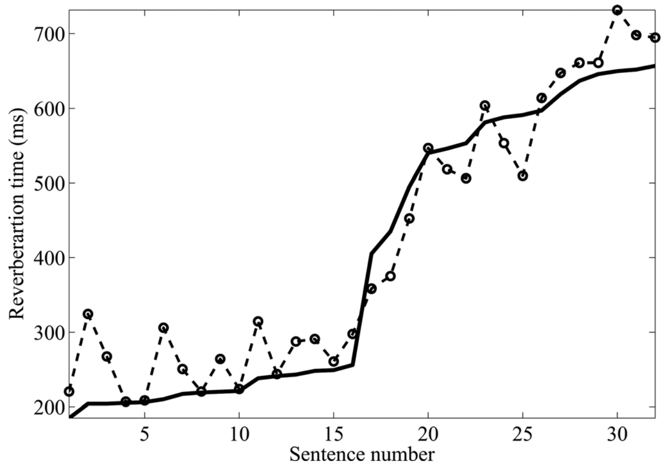

FIG. 3. Estimated RT values using proposed blind (dashed line) and reference non-blind (solid line) methods for all 32 signals in Database B.

Using the training Database A_1, the mapping parameters in Eq. (7) were set to $\alpha = 3.4$ and $\beta = -1170$ ms, in order to minimize the MSE between the estimated RTs using the proposed blind method and the reference non-blind method described in Ref. 15, without affecting the statistical correlation of these two processes. Using this setup, the RT estimates for the entire Database A are depicted in Fig. 2 along with the non-blind RT values, illustrating the overall ability of the proposed algorithm to provide a reliable estimate for a wide RT range.

The RT results for the entire Database B using the proposed algorithm with the same setup as before are shown in Fig. 3, where the statistical correlation in this case achieved the 97% level. The significant increase on this factor can be credited to the reduced reverberation scope covered by Database B in comparison to the additional aspects (three different reverberation setups, wider RT, and RSV ranges, etc.) considered by Database A.

D. Comparison to other approaches

Table IV shows the statistical correlation ρ and the standard deviation σ between the theoretical and estimated T_{60} for both Databases A and B using the algorithms described in Refs. 4 and 8. Table IV also includes results provided by several speech-quality evaluation algorithms, which, in some cases, are closely related to the RT measure.

TABLE IV. Statistical correlation (ρ) and standard deviation (σ) between theoretical and estimated T_{60} for several RT- or quality-estimation algorithms for Databases A and B.

Estimation	Database A		Database B	
algorithm	ρ [%]	σ [ms]	ρ [%]	σ [ms]
Ratnam *et al*. (Ref. 4)	55	254	50	178
Vieira (Ref. 8)	64	234	26	290
R_{DT} (Ref. 12)	58	248	57	190
SRMR (Ref. 13)	72	213	84	105
ITU-T W-PESQ (Ref. 21)	76	197	78	118
ITU-T P.563 (Ref. 22)	29	293	25	198
Proposed algorithm	**91**	**124**	**97**	**46**

This last group of algorithms include, for instance, the reverberation decay time (R_{DT}),[12] the speech-to-reverberation modulation energy ratio (SRMR),[13] and the ITU-T W-PESQ (Ref. 21) and P.563 (Ref. 22) recommendations, all provided by their respective authors for this research. From Table IV, one concludes that the proposed algorithm achieved the highest correlation level and the lowest standard deviation, for both training and testing databases, successfully predicting the RT value in each case.

IV. CONCLUSION

This paper dealt with the RT blind estimation for degraded speech signals. The proposed technique includes four frame-based simple stages, greatly reducing the overall complexity of the resulting approach. Performance of the proposed approach was assessed for two independent databases of reverberant speech, yielding high correlation scores and low standard deviation with respect to estimates provided by a standard non-blind method. Results indicate that the proposed technique can be successfully used to monitor the reverberation effect in practical single-end communications systems.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Wen for making the MARDY database and the reverberation decay time algorithm (Ref. 12) available; Dr. Falk, for providing the SRMR algorithm (Ref. 13); Professor Karjalainen, for providing the non-blind RT estimation algorithm described in Ref. 15; and Dr. Jeub for providing the RIRs given in Ref. 19 for the "natural" reverberation mode of Database A.

[1]D. A. Berkley and J. B. Allen, "Normal listening in typical rooms: The physical and psychophysical correlates of reverberation," in *Acoustical Factors Affecting Hearing Aid Performance*, 2nd ed., edited by G. A. Studebaker and I. Hochberg (Allyn and Bacon, Boston, 1993).
[2]M. R. Schroeder, "New method of measuring reverberation time," J. Acoustic. Soc. Am. **37**(3), 409–412 (1965).
[3]ITU-T Rec. G.191, *Software tools for speech and audio coding standardization* (1995).
[4]R. Ratnam, D. L. Jones, B. C. Wheeler, W. D. O'Brien, Jr., C. R. Lansing, and A. S. Feng, "Blind estimation of reverberation time," J. Acoust. Soc. Am. **114**(5), 2877–2892 (2003).
[5]R. Ratnam, D. L. Jones, and W. D. O'Brien, Jr., "Fast algorithms for blind estimation of reverberation time," IEEE Signal Process. Lett. **11**(6), 537–540 (2004).
[6]J. Vieira, "Automatic estimation of reverberation time," in *Proceedings of the Conv. Audio Engineering Society*, Berlin, Germany (May 2004), pp. 1–7.
[7]N. Xiang, "Evaluation of reverberation times using a nonlinear regression approach," J. Acoust. Soc. Am. **98**, 2112–2121 (1995).
[8]J. Vieira, "Estimation of reverberation time without test signals," in *Proceedings of the Conv. Audio Engineering Society*, Barcelona, Spain (May 2005), pp. 1–7.
[9]M. Wu and D. Wang, "A pitch-based method for the estimation of short reverberation time," Acta. Acust. Acust. **92**, 337–339 (2006).
[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 *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Las Vegas, Nevada (April 2008), pp. 329–332.
[11]H. W. Löllmann and P. Vary, "Estimation of the reverberation time in noisy environments," in *Proceedings of the IEEE International Workshop Acoustic Echo and Noise Control*, Seattle, Washington (September 2008), pp. 1–4.

[12]J. Y. C. Wen and P. A. Naylor, "An evaluation measure for reverberant speech using decay tail modelling," in *Proceedings of the European Signal Processing Conference*, Florence, Italy (September 2006), pp. 329–332.
[13]T. H. Falk and W.-Y. Chan, "A non-intrusive quality measure of dereverberated speech," in *Proceedings of the IEEE International Workshop Acoustic Echo and Noise Control*, Seattle, Washington (September 2008), pp. 1–4.
[14]L. L. Beranek, "Concert hall acoustics—1992," J. Acoust. Soc. Am. **92**(1), 1–39 (1992).
[15]M. Karjalainen, P. Antsalo, A. Mäkivirta, T. Peltonen, and V. Välimäki, "Estimation of modal decay parameters from noisy reponse measurements," in *Proceedings of the Conv. Audio Engineering Society*, Amsterdam, Netherlands (May 2001), pp. 867–878.
[16]ISO Rec.3382, *Measurement of the reverberation time of rooms with reference to other acoustical parameters* (1997).
[17]E. A. P. Habets, "Single-channel speech dereverberation based on spectral subtraction," in *Proceedings of the Workshop Circuits, Systems and Signal Processing*, Veldhoven, Netherlands (November 2004), pp. 250–254.
[18]A. A. de Lima, F. P. Freeland, P. A. A. Esquef, L. W. P. Biscainho, B. C. Bispo, R. A. de Jesus, S. L. Netto, R. Schafer, A. Said, B. Lee, and A. Kalker, "Reverberation assessment in audioband speech signals for telepresence systems," in *Proceedings of the International Conference on Signal Processing in Multimedia Applications*, Porto, Portugal (July 2008), pp. 257–262.
[19]M. Jeub, M. Schäfer, and P. Vary, "A binaural room impulse response database for the evaluation of dereverberation algorithms," in *Proceedings of the International Conference on Digital Signal Processing*, Santorini, Greece (July 2009), pp. 1–4.
[20]J. Y. C. Wen, N. D. Gaubitch, E. A. P. Habets, T. Myatt, and P. A. Naylor, "Evaluation of speech dereverberation algorithms using the MARDY database," in *Proceedings of the IEEE International Workshop Acoustic Echo and Noise Control*, Paris, France (September 2006), pp. 1–4.
[21]ITU-T Rec. P. 862.2, *Wideband Extention to Recommendation, P. 862 for the Assessment of Wideband Telephone Networks and Speech Codecs* (2005).
[22]ITU-T Rec. P.563, *Single-ended Method for Objective Speech Quality Assessment in Narrowband Telephony Applications* (2004).