LOW VARIANCE BLIND ESTIMATION OF THE REVERBERATION TIME

Nicolás López[1,2], Yves Grenier[2], Gaël Richard[2] and Ivan Bourmeyster[1]

[1]Arkamys - 31 rue Pouchet, 75017 Paris, France
[2]Institut Mines-Télécom - Telecom ParisTech - CNRS/LTCI - 37/39 rue Dareau, 75014 Paris, France

ABSTRACT

The reverberation time is a key feature for describing the acoustic properties of a reverberant room. It can be computed from a measured Room Impulse Response but in many applications it has to be estimated blindly. Existing blind methods give accurate estimates but they often exhibit high variance across different speakers. In this paper, a low variance blind estimator of the reverberation time is derived from the decay rate distribution of the signal. The influence of the reverberation time on the statistical moments of the distribution is analyzed and one relevant moment is taken as an estimator. The variance of the estimator is reduced thanks to a prewhitening filter and a modification of the decay rate distribution. Experimental results confirm the accuracy of the method when the observed signal is sufficiently long.

Index Terms— Reverberation time, blind estimation, decay rate distribution, low variance

1. INTRODUCTION

The reverberation time (RT) is one of the main features for describing the acoustics of a room. It is defined as the amount of time required to measure an energy decay of 60 dB after the excitation source is turned off. The RT gives valuable information on the degradation affecting the speech signal [1] and it is needed to calibrate der beration algorithms that are based on a statistical model of the Room Impulse Response (RIR) of the enclosure [2, 3]. It is usually computed from a measured RIR using the well-known Schroeder's backwards integration method [4]. However in a real speech communication context we do not have access to this information and must proceed blindly.

The problem of blind estimation of the RT has largely been addressed in the last decade. Some blind techniques exploit a model of the deformation of the speech signal introduced by reverberation. The RT is then mapped to a measure of the deformation of the temporal [5] or the spectral [6] envelope of the signal. In [7], an Artificial Neural Network is trained to learn reverberation models. Other methods segment the decaying regions of the log-energy envelope of the signal and use linear regression on these regions to track the decay rate [2]. Recently, Maximum Likelihood (ML) approaches have been developed [8, 9]. RT estimates are continuously computed and an order filter is used to choose the most likely value. Wen *et al.* develop a blind method linking the second moment of the decay rate distribution to the RT [10]. The method performs in the Fourier domain. For each analysis frame a linear regression is made on the subband log-energy envelope to compute the decay rates. The method is fast and reliable, but exhibits high variance accross speakers.

In this paper, we introduce a low variance RT estimator based on the decay rate distribution of speech signals. The estimation is performed in the time domain by studying the distribution of the energy ratios between adjacent frames of the energy envelope. The analysis of the relationship between the statistical moments of this distribution and the RT shows that the variance of the negative-side of the distribution is a reliable estimator of the RT. We show in our experiments that a prewhitening stage significantly reduces the variance of the estimator while keeping a small estimation bias. We also show that using a truncated distribution instead of the symmetric one used in [10] improves the accuracy of the estimator.

The paper is organized as follows: in Section 2 we introduce the sound decay model that will be used to derive the RT estimator described in Section 3. The estimator is compared to a state of the art method in Section 4 and some conclusions are suggested in Section 5.

2. MODEL OF SOUND DECAY TAIL

The decay tail of a RIR is often modeled as an exponentially damped Gaussian white noise [11]. Since the RT is defined from a measured RIR, we expect to reliably estimate it in speech segments where the RIR model holds. Thus, we model the decay tail $d(n)$ of speech signals as:

$$d(n) = b(n)\mathrm{e}^{-\delta n} \tag{1}$$

where $b(n) \sim \mathcal{N}(0, \sigma_b^2)$, n is the sample index and δ is the decay rate which is related to the reverberation time by:

$$\delta = 3\ln(10)/RT \tag{2}$$

Using equation (1), we compute the energy envelope of the decay tail, denoted $e(n)$:

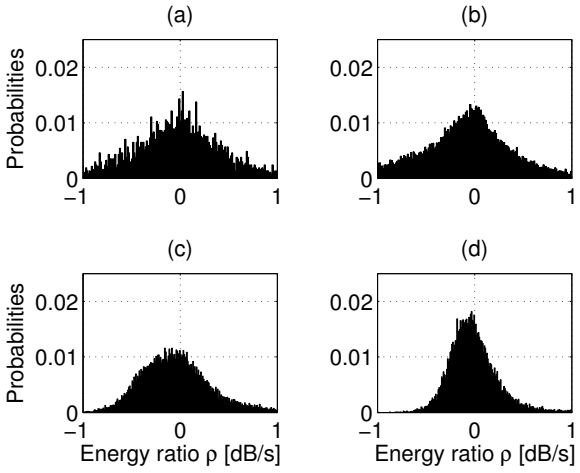

Fig. 1. Distribution of the energy ratios for some RT values: a) Anechoic speech, b) $RT = 250$ ms, c) $RT = 500$ ms and d) $RT = 1$ s.

$$e(n) = \mathrm{E}[d(n)d(n)^*] = \sigma_b^2 \mathrm{e}^{-2\delta n} \quad (3)$$

where $\mathrm{E}[.]$ denotes the expectation operator.

We analyze the temporal energy envelope of the signal with a frame length of N samples and an hop size of R samples. Let's consider the total energy of the envelope $e(n)$ over the m^{th} frame:

$$E_m = \sum_{i=0}^{N-1} e(Rm+i) = \sigma_b^2 \left(\frac{1-e^{-2\delta N}}{1-e^{-2\delta}} \right) \mathrm{e}^{-2\delta R m} \quad (4)$$

Let $\rho(m)$ be the logarithm of the energy ratio between frames m and $m-1$. Using equation (4), a straightforward estimator of the decay rate is obtained:

$$\rho(m) = \log \left(\frac{E_m}{E_{m-1}} \right) = -2\delta R \quad (5)$$

The model assumed in this section holds only if the speech source is a white noise, which is not realistic. For all other speech signals, the decay rate is underestimated. In the next section, we use the distribution of ρ to build an estimator that is robust to modeling errors.

3. BUILDING AN ESTIMATOR

3.1. Energy Decay Distribution

In the following, we store successive estimates of the decay rate in an histogram to get an approximation of the energy decay rate distribution. Fig. 1 shows these histograms for a single speaker and multiple RT. Fig. 1-a) shows the Laplacian nature of the anechoic speech distribution and Figs. 1-b), c) and d) give the shape of the distribution for a RT of 250 ms, 500 ms and 1 s, respectively. As RT increases, negative decay rates appear more often and the distribution becomes asymmetric.

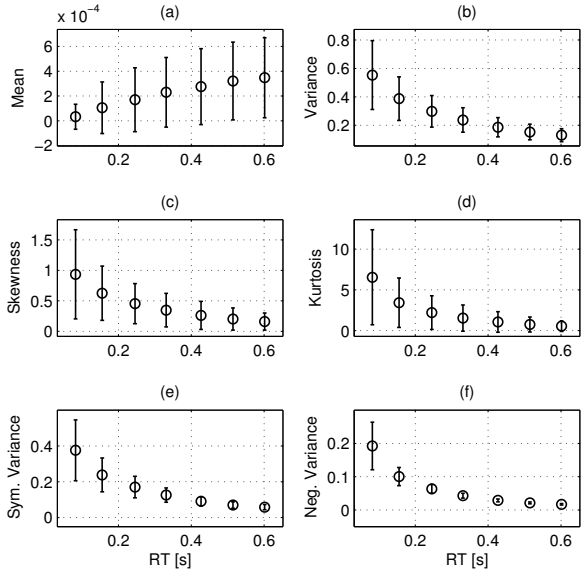

Fig. 2. Statistical moments of the decay rate distribution as a function of RT with 95% confidence intervals represented: a) Mean, b) Variance, c) Skewness, d) Kurtosis, e) Variance of the symmetric distribution [10] and f) Variance of the truncated distribution.

We analyze the average, across several speakers, of the first four statistical moments of the distribution of ρ in order to establish a link between the RT and the shape of the distribution. Figs. 2-a), b), c) and d) respectively show the average mean, variance, skewness and kurtosis of the distribution, as well as the 95% confidence intervals. For those four moments, the confidence intervals are too wide and can lead to confusion if they are used as estimation features. This is explained because the shape of the distribution depends on both the RIR and the considered speaker. In [10], it is assumed that the negative-side of the distribution contains most of the information about the decay rate of the RIR. Thus, the variance of a symmetric distribution having the same negative-side than the original distribution is used in [10] as an estimator for the RT. In this work, we keep this assumption but we study the properties of a truncated distribution where all the positive decay rates of ρ are ignored. Then, we take the variance of the truncated distribution as an estimator. Figs. 2-e) and f) show the variances of the two modified distributions. In both cases the variability due to the speakers is reduced and the lowest confidence intervals are obtained with the variance of the truncated distribution, denoted ν.

Finally, a 3^{rd} order polynomial is fit by least squares in order to determine the decay rate δ as a function of ν:

$$\delta = \alpha_3 \nu^3 + \alpha_2 \nu^2 + \alpha_1 \nu + \alpha_0, \quad (6)$$

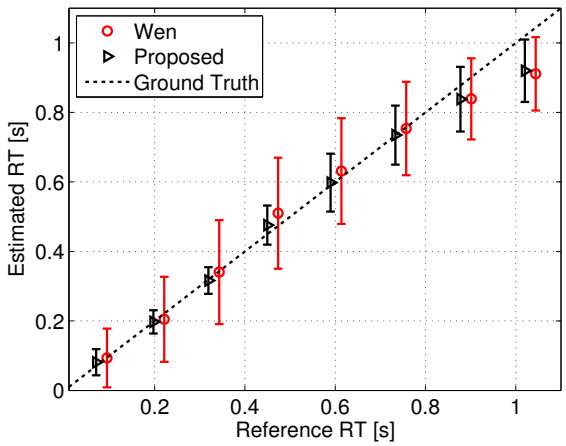

Fig. 3. Bias of the estimator compared to Wen *et al.* method [10]. 95% confidence intervals are shown for some RT values.

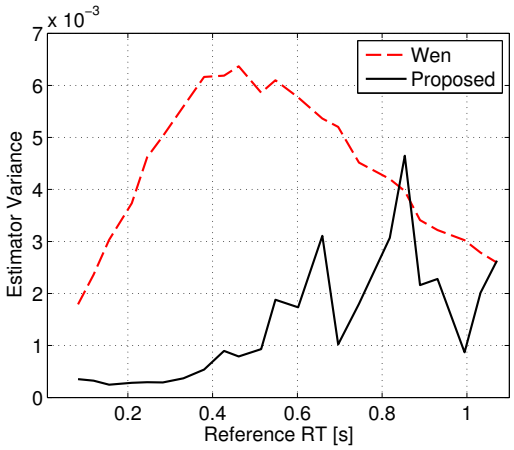

Fig. 4. Variance of: i) Wen *et al.* estimator (dashed), ii) the proposed estimator (solid) for several RT values.

3.2. Prewhitening

The white noise assumption made in Section 2 does not hold for speech. We get closer to the model using a prewhitening filter obtained by linear prediction, for example. According to [12] the LP residual of reverberant speech is explained as the convolution of the LP residual of clean speech by the RIR and, because of this, the effect of reverberation is mainly observable in the LP residual of the signal. We then expect to estimate reliably the RT using the prewhitened signal.

4. EVALUATION

4.1. Settings

For our experiments, we generated 23 RIR with reverberation times from 100 ms to 1100 ms using the Fast Image-Source Method described in [13]. Schroeder's method was then used to validate the reverberation times of the generated RIR. About 7 minutes of anechoic speech from 7 different speakers were taken from the ARCTIC database [14] to train the estimator. For each speaker and each RIR, we pass the signal through a 20^{th} order prewhitening filter with 32 ms long non-overlapping windows. The prewhitened signal is then analyzed in frames of length $N = 64$ ms with 75% overlap. We compute the logarithm of the energy ratios between adjacent frames and build the corresponding histogram. The variance of the negative values of the histogram is then calculated for each speaker and each RIR. Finally, a least squares polynomial fit over the average variance across the speakers gives a function linking the measured variance ν to the decay rate δ, which is linked to the RT by (2). After exhaustive experiments, a 3^{rd} order polynomial was found to be the best fit to the observations with a correlation coefficient of 0.9992.

4.2. Estimator Performance

We test our method in a 7-fold approach. For each fold, we exclude one speaker from the training database and we compute the average, across the remaining speakers, of the variance of the truncated distribution as a function of the RT. The coefficients $\{\alpha_i\}$ of the estimator in (6) are then calculated and stored. Then, we observe about 1 minute of signal from the excluded speaker and compute its decay rate histogram. After taking ν, the variance of the negative-side of the distribution of ρ, we estimate the RT using the learned coefficients. We compare this method to our implementation of the algorithm in [10], where the decay rates are computed by linear fitting on the energy envelope of the signal and the histogram is computed over exactly the same 1-minute long observation as above. The average estimated RT and the variance of the estimator are analyzed for both methods.

In Fig. 3, we show the estimated RT as well as the 95% confidence intervals. Both algorithms give an unbiased estimation for low RT and a slight bias as RT increases. We analyze the variance of both estimators in Fig. 4. The proposed method has a very low variance for RT smaller than 500 ms. For higher RT, the variance increases with the RT but it remains lower than the one of Wen *et al.* method for almost every RT. This ensures better robustness to speaker variability. This gain in accuracy is obtained at the cost of longer computation as we need to observe at least 1 minute of speech signal to converge to a correct estimate. This is not necessarily unpractical because in some situations, such as in-car dereverberation, the acoustic environment is quite stable and allows for an incremental learning of the properties of the enclosure.

We finally study the influence of the prewhitening stage by ignoring it in the same training and testing scheme described above. For completeness, we justify the use of the truncated decay rate distribution by comparing our RT esti-

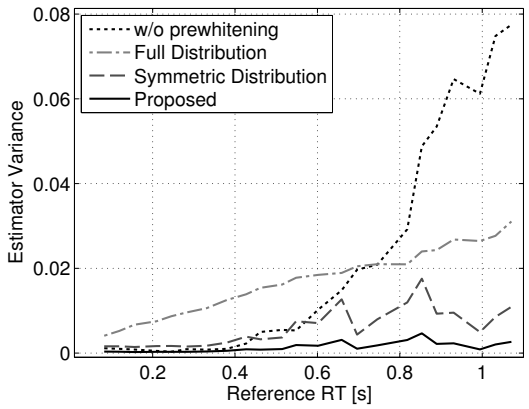

Fig. 5. Variance of the estimator when the prewhitening stage is ignored (dotted), when the whole decay rate distribution is considered (dash-dotted), when the symmetric distribution is used (dashed) and using the truncated distribution and the prewhitening stage as described above (solid).

mator to the one obtained when the whole distribution or the symmetric distribution of the prewhitened signal are considered. We analyze the variance of the estimates for each case in Fig. 5. It is clear that both the prewhitening and the truncation of the distribution reduce the variance of the estimates, particularly for high reverberation times. This is because both modifications narrow down the gap between the signal and the model given in Section 2.

5. CONCLUSION

We presented a robust method for the blind estimation of the reverberation time of a room. The decay rates of the time domain energy envelope of the signal are computed. We analyzed the influence of the reverberation time on the shape of the distribution of the decay rates. The variance of the negative-side of the distribution was found to be a good feature for the estimation and it was linked to the reverberation time using a 3^{rd} order polynomial. We showed that the use of a prewhitening filter as a preprocessing stage helps to reduce the variance of the estimator. Our evaluation showed that, if the analyzed signal is sufficiently long, the proposed estimator reliably estimates the reverberation time while having little dependency on the considered speaker.

6. REFERENCES

[1] J.B. Allen, "Effects of small room reverberation on subjective preference," *J. Acoust. Soc. Am.*, vol. 71, pp. S5, 1982.

[2] K. Lebart, J.M. Boucher, and P.N. Denbigh, "A new method based on spectral subtraction for speech dereverberation," *Acta Acustica*, vol. 87, no. 3, pp. 359–366, 2001.

[3] E.A.P. Habets, *Single- and Multi-Microphone Speech Dereverberation using Spectral Enhancement*, Ph.D. thesis, Technische Universiteit Eindhoven, Eindhoven, The Netherlands, June 2007.

[4] M.R. Schroeder, "New method of measuring reverberation time," *J. Acoust. Soc. Am*, vol. 37, no. 3, pp. 409–412, 1965.

[5] L. Couvreur, C. Ris, and C. Couvreur, "Model-based blind estimation of reverberation time: Application to robust asr in reverberant environments," in *7th European Conference on Speech Communication and Technology*, Aalborg, Denmark, 2001.

[6] M. Wu and D.L. Wang, "A pitch-based method for the estimation of short reverberation time," *Acta Acustica*, vol. 92, no. 2, pp. 337–339, 2006.

[7] T.J. Cox, F. Li, and P. Darlington, "Extracting room reverberation time from speech using artificial neural networks," *J. Audio Eng. Soc.*, vol. 49, no. 4, pp. 219–230, 2001.

[8] R. Ratnam, D.L. Jones, B.C. Wheeler, W.D. OBrien Jr, C.R. Lansing, and A.S. Feng, "Blind estimation of reverberation time," *J. Acoust. Soc. Am.*, vol. 114, pp. 2877–2892, 2003.

[9] H.W. Löllmann, E. Yilmaz, M. Jeub, and P. Vary, "An improved algorithm for blind reverberation time estimation," in *Proc. IWAENC*, Tel Aviv, Israel, 2010.

[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. ICASSP*, Las Vegas, USA, 2008, pp. 329–332.

[11] J.D. Polack, *La Transmission de l'énergie sonore dans les salles*, Ph.D. thesis, Université du Maine, Le Mans, France, 1988.

[12] N.D. Gaubitch, Thomas M.R.P., and P.A. Naylor, "Dereverberation using lpc based approaches," in *Speech Dereverberation*, P.A. Naylor and N.D. Gaubitch, Eds., pp. 99–105. Springer, 2010.

[13] E.A. Lehmann and A.M. Johansson, "Diffuse reverberation model for efficient image-source simulation of room impulse responses," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, no. 6, pp. 1429–1439, 2010.

[14] J. Kominek and A.W. Black, "The cmu arctic speech databases," in *5th ISCA Workshop on Speech Synthesis*, Pittsburgh, USA, 2004.