AUTOMATIC ESTIMATION OF REVERBERATION TIME FROM BINAURAL SIGNALS

Sampo Vesa[*]

Helsinki University of Technology
Telecom. Software and Multimedia Lab.
P.O. Box 5400, FIN-02015, HUT, Finland
svesa@tml.hut.fi

Aki Härmä[†]

Helsinki University of Technology,
Lab. of Acoustics and Audio Sig. Proc.
P.O. Box 3000, FIN-02015 HUT, Finland

ABSTRACT

An estimate of the reverberation time (RT) at the space of usage can be useful in many communications applications, such as augmented reality audio and intelligent hearing aid devices. This paper presents a method to measure the reverberation time from two microphone signals. The analysis is based on locating suitable sound segments for RT analysis by using short-time energy and inter-channel coherence measures, followed by the Schroeder integration method, line fitting and finally statistical analysis. The line fitting is used to estimate the slope of the decay. In this paper, we propose a method where the slope is estimated in the region that maximizes the correlation coefficient of the least squares method. This makes the estimation results more accurate than if fixed limits, e.g., -5 to -25 dB on the decay curve, were used, thanks to the absence of the systematic error caused by bending of the decay curves. The system performance was evaluated using a real-time version of the algorithm.

1. INTRODUCTION

In many wearable and mobile audio applications it would be advantageous to adjust the performance of the system as a function of the reverberation time in the space around the user. For example, in the mobile augmented reality audio (MARA) system introduced in [1] an estimate of the reverberation time in the environment is used to control the synthesis of virtual sound events. An estimate of the acoustical properties of the user's environment is also a very useful cue for sensing the context of a mobile user.

Typically no *a priori* information about the reverberation time (RT) is available. It is also not practical to measure the RT using, e.g., a controlled excitation signal such as a noise burst or an impulse. Thus the sounds naturally present in the environment have be used to determine the reverberation time of the surrounding acoustic space. This paper presents one such method, which uses specifically binaural signals to estimate the reverberation time. The binaural signals are microphone signals recorded close to the two ears of the user. The nature of the binaural input signal is taken advantage of in the presented method.

Generally, the problem is to find the decay of the room impulse response function $h(n)$ from an observed signal which is a convolution between an unknown input signal $x(n)$ and the room response $y(n) = \sum_{k=0}^{\infty} h(k)x(n-k)$. Both $h(n)$ and $x(n)$ are unknown and therefore cannot be separated uniquely from a single observation. However, we can safely assume that the decay of $h(n)$ is relatively stationary compared to the variation in observed signal $x(n)$. Therefore we may try to locate specific regions, e.g., transients and rapid offsets of source signals, where a useful estimate of the decay can be estimated from the decay in the observed signal. There are basically two types of approaches for automatic RT estimation mentioned in literature. Some approaches use some sort of a segmentation procedure to find the interesting sounds from a continuous signal, and perform the RT analysis on those parts of the signal only. Another class of methods perform calculations on the signal continuously, regardless of the signal content. These methods are sometimes termed *blind* methods, in analogy with the blind source separation (BSS) problem.

The former group of methods includes using the decay of the envelope of the autocorrelation function [2], neural networks trained on artificial impulse responses with different reverberation times [3] and automatical detection of decaying segments followed by Schroeder integration and linear regression [4] [5] [6] (Schroeder integration is not mentioned in [4]). The latter group includes maximum likelihood estimation based methods [7] [8] [9] and blind deconvolution, which gives the impulse response as a byproduct. However, blind deconvolution only works when the impulse response is minimum phase, a condition that is not fulfilled in most real acoustic spaces [7].

The method presented in this paper is a partially blind method based on locating segments in the input signal which are suitable for the estimation of the reverberation time. The algorithm is partially based on the analysis of the short-term coherence between the two input channels. In addition, we propose a modification of the classical line fitting technique which gives more accurate RT estimates in the present application. The proposed algorithm has been developed to a real-time prototype running in an ordinary Linux workstation.

2. DESCRIPTION OF THE METHOD

This section presents the building blocks of the RT estimation method in the order that they are ran. The algorithm takes a binaural two-channel signal as an input. The output is a new reverberation time estimate for each segment.

2.1. Segmentation

The parts of the sound signal that are potential candidates for reverberation time estimation are detected using a simple activity

[*]Mr. Vesa's work was supported by the Nokia Research Center.
[†]Currently with Philips Research, Eindhoven, The Netherlands.

detector based on the short-time signal energy. The signal energy is calculated from short frames and an estimate for the background noise level is formed by calculating the average of number of previous frame energies. If the energy of the current frame exceeds that of background noise by a certain amount (e.g., 10 dB), a start point of a sound event is detected and the algorithm starts recording the input frames into a larger buffer that will contain the entire event (segment). This continues for a few seconds or until the energy of the current frame falls below the noise level (plus some constant, which could be zero).

When the event ends, the entire buffer of the samples of the two channels is passed on to further processing. In the real-time implementation all further processing is done in a separate thread with lower priority to avoid overloading the real-time thread.

2.2. Determination of the integration limits

In practical applications of the Schroeder method, the upper limit of integration (T_i in Eq. 1) has to be decided. According to [10] and [11], the choice of T_i should be made so that T_i is close to the point where the decaying signal dives below the noise floor. A special method for locating the upper limit has been proposed in [12]. The location in time up to which the decay curve is calculated (T_d) should also be decided.

The upper limit of integration T_i is located by inspecting a smoothed logarithmic energy envelope of the segment and using the latest noise level estimate do decide where the decay curve dives below the noise level. The decay curve starting point T_d, which is ideally the point where the direct sound and early reflections have just ended and the diffuse decay has started, is located based on short-time coherence between channels. The average of short-time coherence has been previously used in hearing-aid algorithms to determine how diffuse the sound field around the user is [13]. In this work the short-time coherence function is evaluated from overlapping frames of the recorded segment. The start of diffuse sound T_d is located by thresholding the average coherence. A high coherence value corresponds to direct sound and reflections. Thus the algorithm counts the number of average coherence values that are over a certain threshold and converts that to number of samples. That number is added to the sample index of the envelope peak value to get an estimate for T_d. Naturally, the peak value might be at any point during the coherent sound, so T_d is always more or less overestimated this way.

2.3. Testing of the segments

Each segment is subjected to three tests in order to determine if the segment is suitable for reverberation time analysis. Each test involves calculating a certain figure from the segment data and checking whether the value is within limits.

The first test fits an optimal line to the energy envelope using the least squares method and thresholds the correlation coefficient of line fit. The fit is done from the peak of the envelope to buffer end. If the correlation coefficient is too low, the segment is discarded. The purpose of this test is to discard segments that have energy envelopes that are very far from being close to linear.

The maximum value of the average of short-time coherence (calculated when locating the integration limits, see Section 2.2) is thresholded to check that the sound segment is a transient one. This is based on the fact that transient sounds gives rise to a sharp peak in the average of the short-time time coherence at the location of the sound onset.

Finally, the spectral centroid is calculated from a single FFT window starting from the index of the maximum value of the envelope. The centroid value will be checked to be within a certain band, e.g. 500-5000 Hz, to avoid sounds with frequencies concentrated too low or too high from biasing the estimation results.

2.4. Schroeder integration

The decay curve is calculated using the standard Schroeder method [14]. A practical formula for applying the method is [10]

$$D(t) = N \int_t^{T_i} h^2(\tau) d\tau \qquad (1)$$

where N is a constant proportional to the PSD of the noise on the frequency range measured and T_i is the upper limit of integration. In this work N is set to one and the decay curve is scaled so that the maximum value, i.e., the value of the decay curve at T_i, is zero decibels. The decay curve is evaluated starting from T_i up to point T_d, where the diffuse decay starts.

2.5. Least squares fit with fixed or variable range

The limits of the least squares fit are important for accurateness of the RT estimates. Even a small bending of the integration curve may strongly affect the slope of the fitted line and the RT calculated based on the slope. To extract the RT value from the decay curve, a standard least squares fit is applied to the decay curve. The simplest idea is to choose the line fitting range to be fixed, e.g. -5 to -35 dB (T_{30}) or -5 to -25 dB (T_{20}). In this algorithm it is also possible to set the rightmost limit of line fitting to maximize the correlation coefficient of the line fitting. This is hypothesized to lead to more accurate RT estimates, even though the differences between different types of RTs, e.g. T_{20} and T_{30}, are forgotten (because of this, reverberation time will be denoted by RT from now on, instead of T_{30} or T_{60}).

The upper panel of Figure 1 presents an example decay curve with fixed line fitting. As can be seen, the fit from -5 to -35 dB slightly overestimates the slope of the decay curve. The middle panel of Figure 1 shows the RT and the lower panel plots the correlation coefficient r^2, both as a function of the rightmost line fitting limit.

The true RT of the room was approximately 0.8 seconds, so if the rightmost limit is at -35 dB, the RT will be underestimated by about 0.1 seconds in this case. By picking the RT from the location of maximum correlation coefficient, the RT will be much closer to the true value, as can be verified from Figure 1.

It is difficult to acquire a reliable estimate of the room RT from a single acoustic event. Therefore, the last part of an automatic reverberation time estimation algorithm is statistical analysis of several consecutive estimates, in this case all estimates up to the current point in time. Different tools could be used, such as the mean, the median or order statistics. In this article we use a method where we pick the (first) maximum of the RT histogram [6]. In this algorithm, the bin center of the first maximum is used. The histogram bins are all of width 0.58 seconds (25 adjacent bins between 0.05 to 1.5 seconds).

3. EVALUATION

The algorithm was tested with real binaural recordings made in an office room where the measured RT is approximately 0.8 seconds.

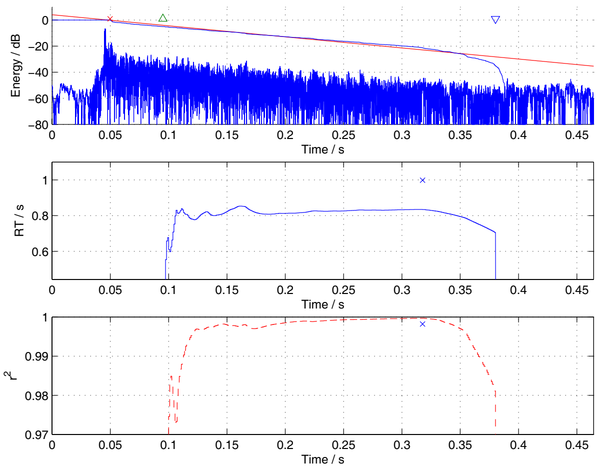


Fig. 1. Example of finding least squares fit range. The line in the upper panel is fit to a range from -5 to -35 dB on the decay curve (denoted by $\triangle$ and $\bigtriangledown$). The starting point for the decay curve (T_d) in the upper panel and the correlation coefficient maximum in the middle and lower panels are marked by 'x'.

The recording consisted of 2 minutes of impulsive sounds, e.g. hand claps, and some miscellaneous sounds.

Figure 2 presents all 38 RT estimates as a function of estimate index for both fixed (-5 to -25 dB) and variable line fitting limits. In the case of variable limits (right panel), the estimates are centered around the true value (0.8 s, denoted by dashed line). This is verified by the means calculated over all estimates, $\mu_{fix} = 0.66$ s and $\mu_{var} = 0.83$, the latter of which is much closer to the true value. The standard deviations are almost equal in both cases ($\sigma_{fix} = 0.20$ s and $\sigma_{var} = 0.15$ s). From the final histograms of all estimates in Figure 5 it is also clear that for the variable limits algorithm the distribution is centered closer to the the true value than in the fixed limits case. It seems that in the fixed limits case there is a systematic bias that is removed when the line fitting limits are chosen to maximize the correlation coefficient.

The evolutions of three different statistics (the mean, median and histogram peak) as a function of estimate index are presented in Figures 3 and 4, for fixed and variable line fitting limits, respectively. The ± 100 ms range around the true value is denoted by dotted lines. The variable limits case shows better performance with the mean and the median, both of which are well within the ± 100 ms range during the entire algorithm run. Picking the histogram peak seems to have comparable performance in both cases and it seems to show similar behavior, resulting in underestimation of the RT.

One typical application of the proposed algorithm is to control the synthesis of virtual audio events such that they have a reverberation subjectively similar to the true acoustic environment around the user. At the moment, it is not known if the obtained RT estimate is always sufficiently accurate for that. This is a part of the future work of the authors. Informal listening experiments with the real-time system have given promising results. It is known that the hearing mechanism is not particularly sensitive to changes in the reverberation time. For example, a classical result by Seraphim [15] suggests that the just noticeable difference in reverberation time, between 0.5 and 2 s, is 40 %. In the present experiment in the room with the reverberation time of 0.8 s the standard deviation of the error was less than 25 %, which, in this respect, is an encouraging result.

Another potential application is to use the algorithm as a part of a system for automatic sensing of the context of the user. In this application the errors in the range of 0.1-0.2 s are probably not significant. It is usually sufficient that the algorithm can discriminate a few cases such as when the user is outdoors (nearly free field), in a small damped room, a large reverberant room such as an auditorium, or in a highly reverberant space.

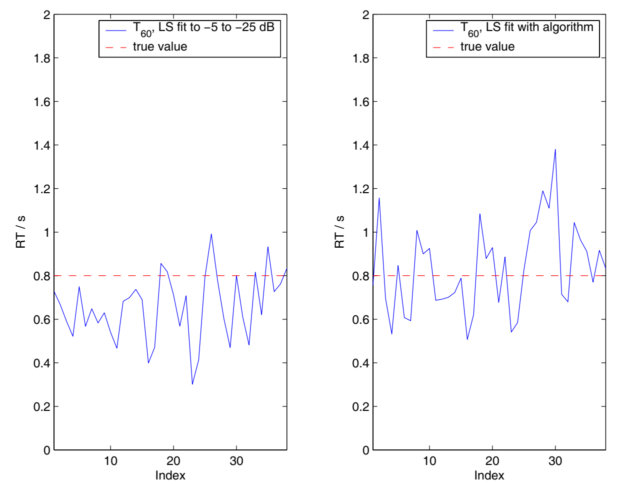


Fig. 2. Estimates of RT for room A152 with and without variable line fitting range.

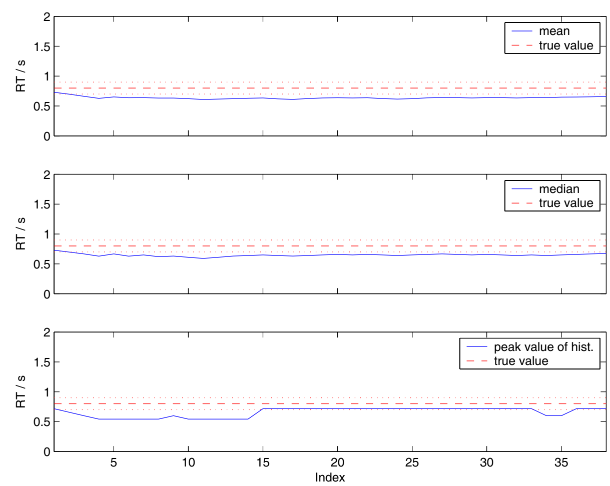


Fig. 3. Three different statistics calculated from RT estimates for room A152, fixed line fitting range -5 to -25 dB.

4. CONCLUSIONS

An algorithm for estimating the reverberation time from binaural signals was presented in this paper. The novel feature of the algorithm is the use of variable line fitting limits for estimating the

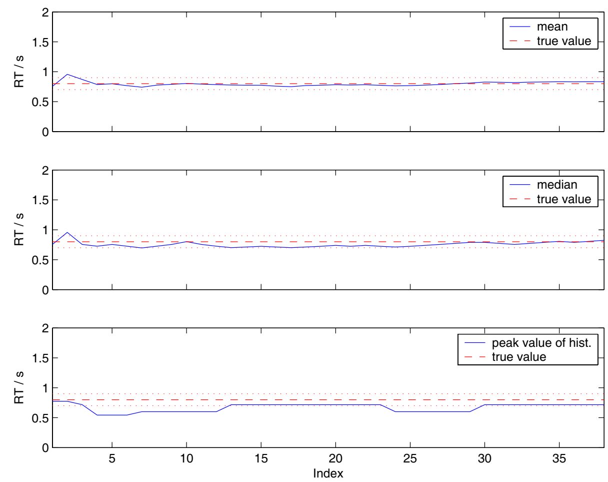
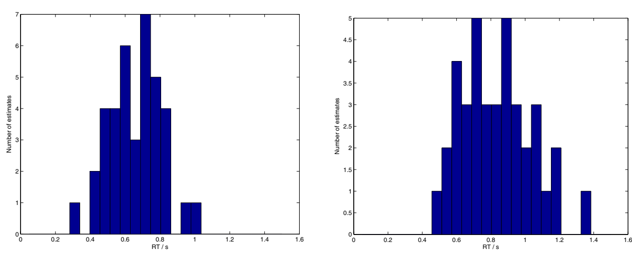


Fig. 4. Three different statistics calculated from RT estimates for room A152, variable line fitting limits.



(a) Fixed limits (-5 to -25 dB) (b) Variable limits

Fig. 5. Histograms of T_{60} estimates for room A152.

slope of the decay curves. The results show that variable line fitting limits results in more accurate estimates, compared to keeping the limits fixed.

It can be anticipated that the accuracy of the proposed algorithm is sufficient for its typical applications. Those are the control of the auralization module in a mobile augmented reality audio system [1], and the sensing of the acoustic context of the user. However, more tests should be done with various reverberation times to verify the validity of the approach.

5. ACKNOWLEDGMENT

This work was conducted in the KAMARA (Killer Applications for Mobile Augmented Reality Audio) project in the Telecommunications and Multimedia Laboratory (TML) of Helsinki University of Technology (HUT). The project was funded by Nokia.

6. REFERENCES

[1] A. Härmä, J. Jakka, M. Tikander, M. Karjalainen, T. Lokki, J. Hiipakka, and G. Lorho, "Augmented Reality Audio for Mobile and Wearable Appliances," *Journal of The Audio Engineering Society*, vol. 52, no. 6, pp. 618–639, June 2004.

[2] M. Hansen, "A Method for Calculating Reverberation Time from Musical Signals," Tech. Rep. 60, The Acoustics Laboratory, Technical University of Denmark, Building 352, DK-2800 Lynbgy, 1995.

[3] T. J. Cox and F. F. Li nand P. Darlington, "Extracting Room Reverberation Time from Speech Using Artificial Neural Networks," *Journal of The Audio Engineering Society*, vol. 49, no. 4, pp. 219–230, April 2001.

[4] K. Lebart, J.-M. Boucher, and P. Denbigh, "A New Method Based on Spectral Subtraction for Speech Dereverberation," *Acustica/Acta Acustica*, vol. 87, no. 3, pp. 359–366, 2001.

[5] A. Baskind and O. Warusfel, "Methods for Blind Computational Estimation of Perceptual Attributes of Room Acoustics," in *Proceedings of the AES 22nd International Conference on Virtual, Synthetic and Entertainment Audio (AES22)*, Espoo, Finland, June 2002.

[6] J. Vieira, "Automatic Estimation of Reverberation Time," in *Proceedings of the AES 116th International Convention*, Berlin, Germany, May 2004.

[7] 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," *Journal of The Acoustical Society of America*, vol. 114, no. 5, pp. 2877–2892, November 2003.

[8] R. Ratnam, D. L. Jones, and W. D. O'Brien Jr., "Fast Algorithms for Blind Estimation of Reverberation Time," *IEEE Signal Processing Letters*, vol. 11, no. 6, pp. 537–540, 2004.

[9] L. Couvreur and C. Couvreur, "Robust Automatic Speech Recognition in Reverberant Environments by Model Selection," in *Proceedings of the International Workshop on Hands-Free Speech Communication (HSC-2001)*, Kyoto, Japan, April 2001, pp. 147–150.

[10] W. T. Chu, "Comparison of Reverberation Measurements Using Schroeder's Impulse Method and Decay-Curve Averaging Method," *Journal of The Acoustical Society of America*, vol. 63, no. 5, pp. 1444–1450, 1978.

[11] D. R. Morgan, "A Parametric Error Analysis of the Backward Integration Method for Reverberation Time Estimation," *Journal of The Acoustical Society of America*, vol. 101, no. 5, pp. 2686–2693, 1997.

[12] A. Lundeby, T. E. Vigran, H. Bietz, and M. Vorländer, "Uncertainties of Measurements in Room Acoustics," *Acustica*, vol. 81, pp. 344–355, 1995, Dedicated to Prof. Dr. Heinrich Kuttruff on the occasion of his 65th birthday.

[13] T. Wittkopp, *Two-Channel Noise Reduction Algorithms Motivated by Models of Binaural Interaction*, Ph.D. thesis, Carl von Ossietzky University Oldenburg, March 2001.

[14] M. R. Schroeder, "A New Method of Measuring Reverberation Time," *Journal of The Acoustical Society of America*, vol. 37, pp. 490–412, 1965.

[15] H.-P. Seraphim, "Untersuchungen über die Unterschiedsschwelle exponentiellen abklingens von Rauschbandimpulsen," *Acustica*, vol. 8, pp. 280–284, 1958.