Estimation of Reverberation Time from Binaural Signals Without Using Controlled Excitation

Sampo Vesa

Master's Thesis presentation on 22nd of September, 2004

Outline

Background

The problem

The algorithm

Evaluation results

Future work and conclusions

Background

Motivation and goals of the work

- An RT estimate would be beneficial in many applications
- It is not feasible to feed a measurement signal into the environment
- Passively received binaural signal is available in some applications
- The goal of this work was to develop a reverberation time estimation method that takes advantage of the binaural nature of the signals

The problem

Estimation of reverberation time from a systems theory perspective

- The reverberation time (RT) is a property of an acoustic space, having impulse response $h(n)$
- Only the output $y(n)$ of the system is observed:
  $y(n) = \sum_{k=0}^{\infty} h(k)x(n-k)$
- Estimate the decay of $h(n)$ by observing $y(n)$ only
- If $h(n)$ is regarded as stationary and $x(n)$ as time varying, certain parts of $y(n)$ can be used for estimating the decay (transients and rapid offsets)
- The approach chosen for this work: detect such parts of the signal and perform RT analysis on those segments only

The problem

Previous approaches

- A rough division of the methods into two categories:
    1. *Blind methods* do not make any assumptions of the signal, e.g. maximum likelihood estimation based methods [8] [3]
    2. *Partially blind methods* use prior information about the signal and usually have some sort of a segmentation procedure, e.g. autocorrelation length of musical signals [5], neural networks [4], locating decaying segments followed by backwards integration and/or line fitting [6] [1] [9]
- The method presented in this work falls into the latter category

The algorithm

Structure of the proposed algorithm

1. Segmentation
2. Locating the limits of Schroeder integration
3. Testing the segments
4. Backwards integration (if segment was accepted)
5. LS fit with fixed or variable range $\longrightarrow$ RT estimate
6. Statistical analysis on all RT values up to this point $\longrightarrow$ final RT estimate

The algorithm

Segmentation

- *Coarse segmentation* detects interesting sound events based on short-time energy of the signal
- The detection of events is based on energy difference thresholding
- An estimate for the background noise level is continuously calculated and a large enough sudden deviation results in a detected event

Finding the limits of Schroeder integration

- A practical formula for applying the Schroeder method is [2]:

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

- *Fine segmentation* attempts to find optimal Schroeder integration limits:
  - T_i is the upper limit of integration in Eq. 1
  - T_d is the point up to which the decay curve is evaluated

The algorithm

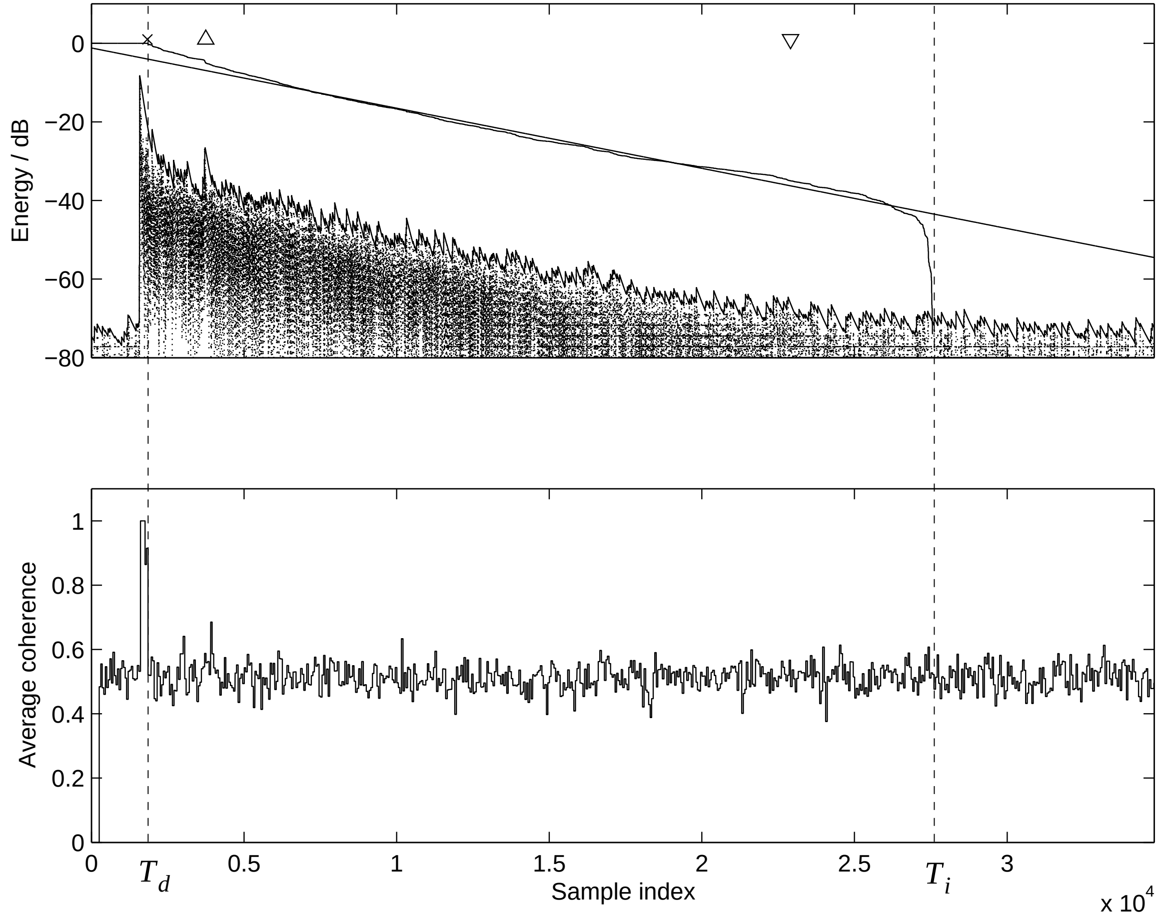

Figure 1: An example of Schroeder integration with the limits T_i and T_d

The algorithm

Finding T_i, the upper limit of Schroeder integration

- T_i should ideally be at the point where the decay “dives” into the noise floor
- A special algorithm for locating T_i is reported in [7]
- This work uses a simpler approach based on calculating a probability density function estimate from an energy envelope of the segment
- Details can be found from the thesis

The algorithm

Finding T_d, the point up to which the decay curve is evaluated

- T_d should ideally be at the point where the diffuse decay starts
- The short-time average interaural coherence (STAIC) has been previously used for measuring the diffusiveness of an acoustical situation [10]
- The STAIC is evaluated from short-time Fourier transforms
- Calculate the length of the part of the segment that has STAIC values over a certain threshold (e.g. 0.8) and sum with the location of the maximum of the envelope
- Always more or less overestimated this way (does not matter)
- A simpler alternative: locate the -5 dB point on the envelope

Testing the segment

- Three tests are performed for each segment to decide whether the segment is suitable for RT analysis
  1. If the energy-time curve is not linear enough (on dB scale), the segment should be discarded $\longrightarrow$ test the linearity of the envelope by least squares fit and thresholding the correlation coefficient
  2. Transient sounds are the best for RT analysis $\longrightarrow$ test transience by thresholding the maximum of the STAIC calculated in the previous step
  3. RT varies as a function of frequency, the sounds used for RT analysis should have frequencies concentrated in the middle $\longrightarrow$ calculate the spectral centroid and require the value to be in a certain range (say, 500-5000 Hz)

The algorithm

Backwards integration (the Schroeder method)

- If the segment passed all three tests, the decay curve is calculated for range $[T_i,\ T_d]$ by using discretized version of the Schroeder method
- Eq. 1 is the basis of this section of the algorithm

The algorithm

Line fitting with fixed or variable limits

- Least squares method is used to fit a line to the decay curve
- RT easily derived from the slope of the line
- Normally the line is fit to a range of -5 to -35 dB (T_{30}) or -5 to -25 dB (T_{20})
- The signal-to-noise ratio (SNR) does not always permit this
- Solution: fit the line to a range that maximizes the correlation coefficient
- Removes the possible systematic bias caused by bending of the decay curves!

The algorithm

Perform statistical analysis

- Finally, statistical analysis is performed on all estimates including the current one
- Possible statistics to use: mean, median, order statistics, peak of histogram...
- The first peak of the histogram sounds good for this application
- Three different statistics (mean, median and histogram peak) were compared in the evaluation part of the thesis

The algorithm

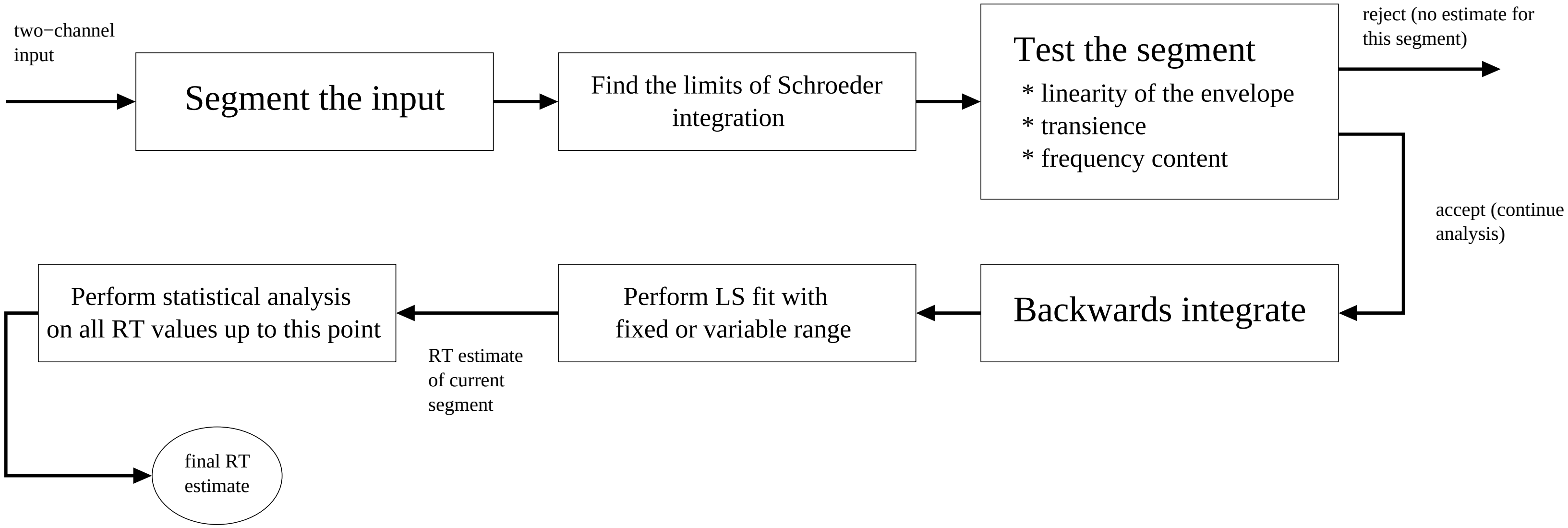

Figure 2: Flowchart of the algorithm

Evaluation results

Testing the algorithm

- Real-world binaural recordings from two different spaces were used to test the algorithm performance
- The work room of the author (A152) has measured RT of ≈ 0.8 s
- The lecture hall T3 has measured RT of ≈ 0.6 s
- The recordings consisted of miscellaneous sounds, hand claps and other impulsive sounds

Evaluation results

Evaluation results

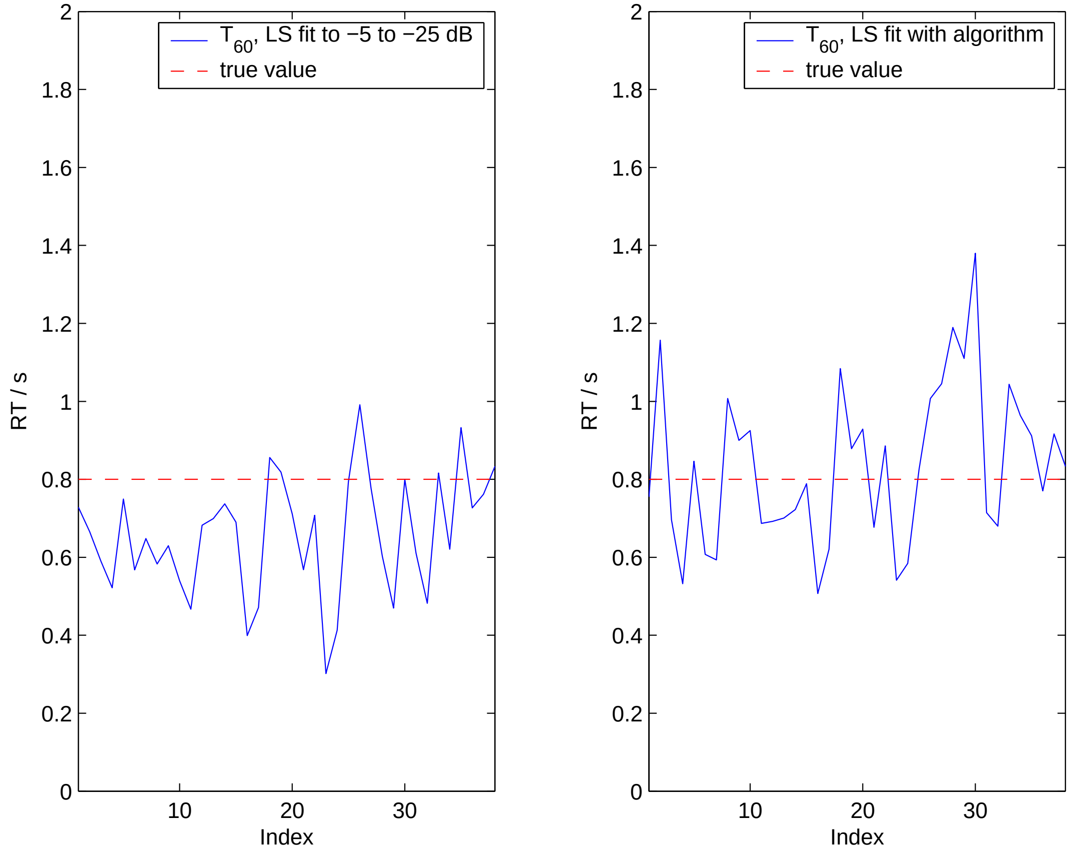

Figure 3: Estimates of T_{60} for room A152 with and without least squares limit lookup, real recording

Evaluation results

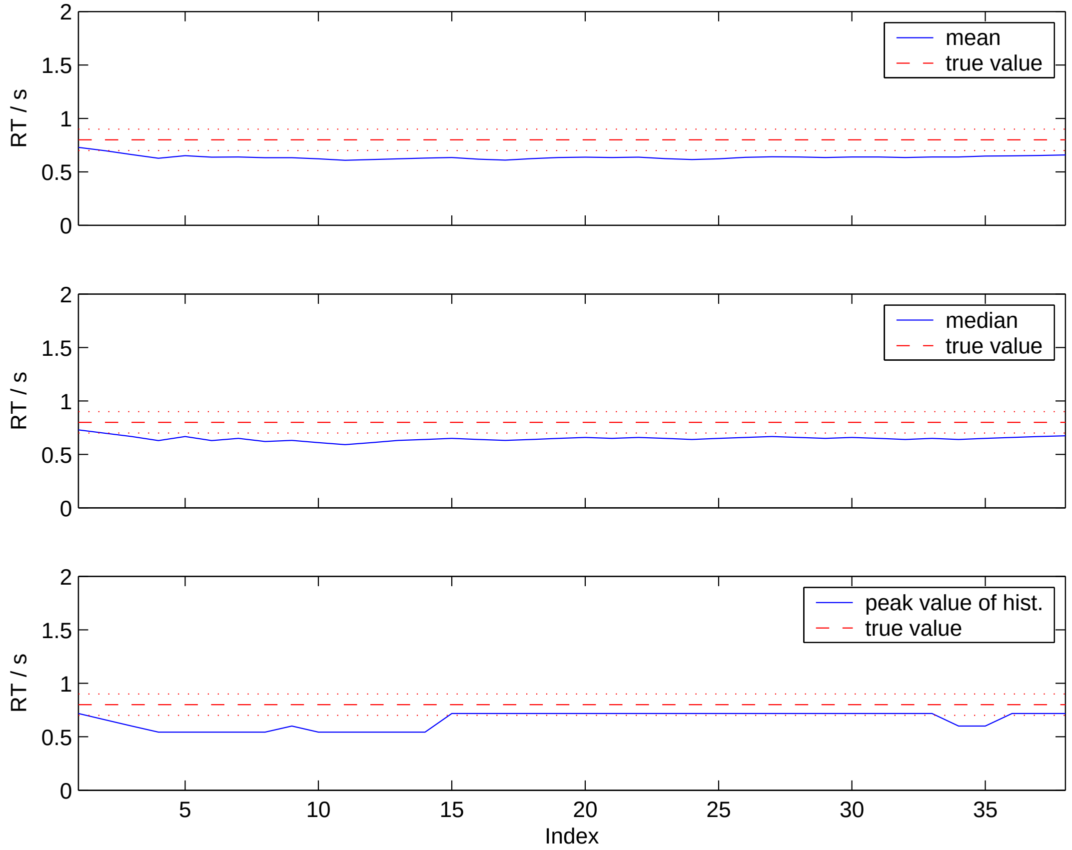

Figure 4: Three different statistics calculated from T_{60} estimates for room A152, real recording, line fitting range -5 to -25 dB

Evaluation results

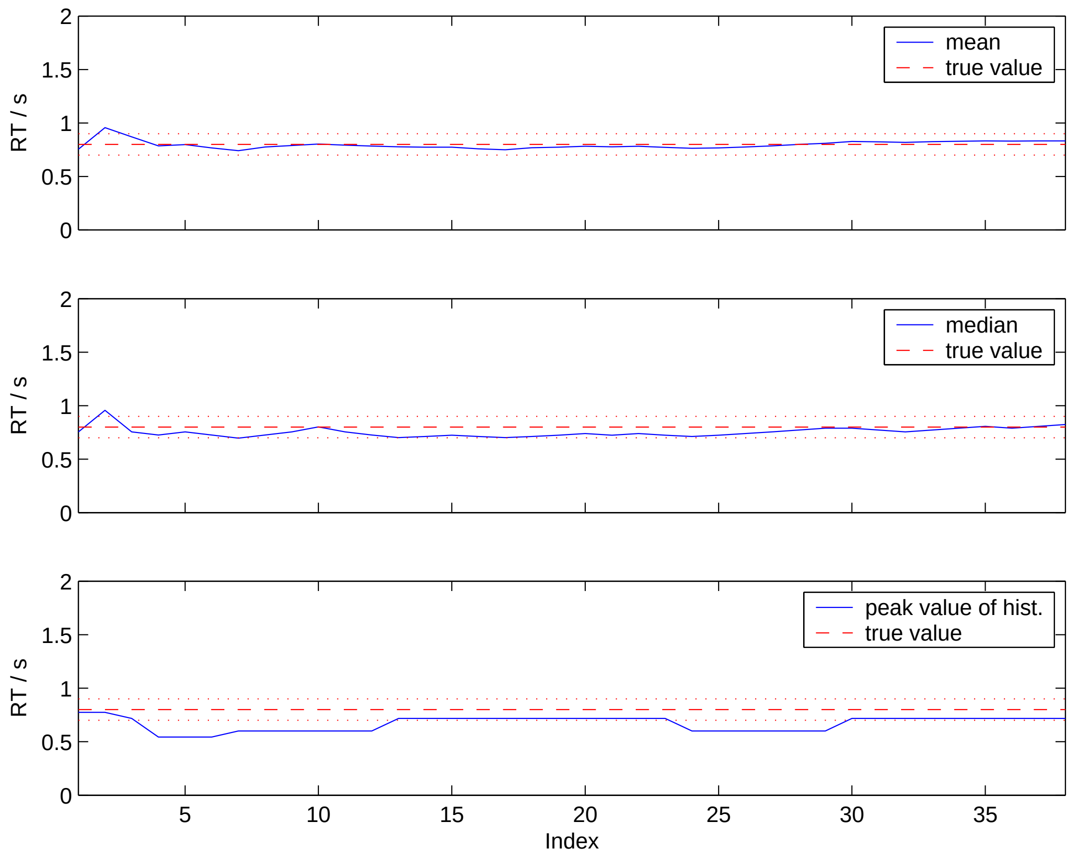

Figure 5: Three different statistics calculated from T_{60} estimates for room A152, real recording, variable line fitting limits

Evaluation results

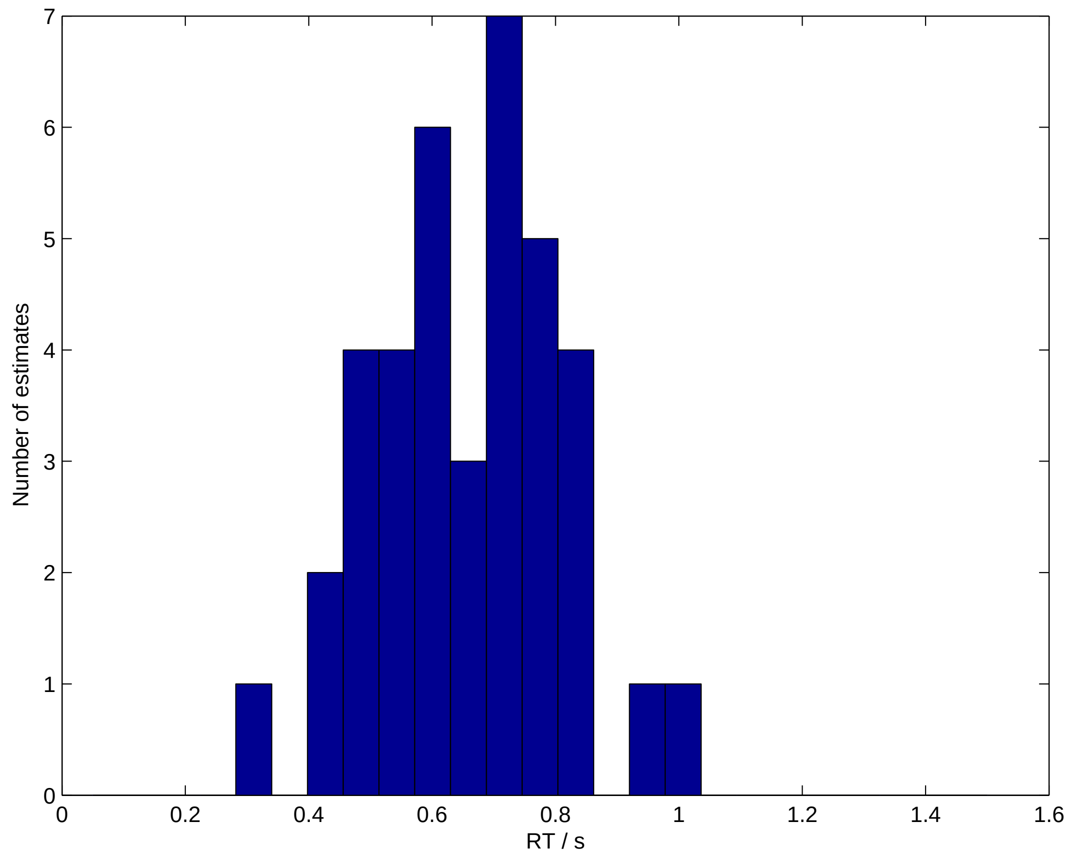

Figure 6: Histogram of T_{60} estimates for room A152, real recording, line fitting range -5 to -25 dB

Evaluation results

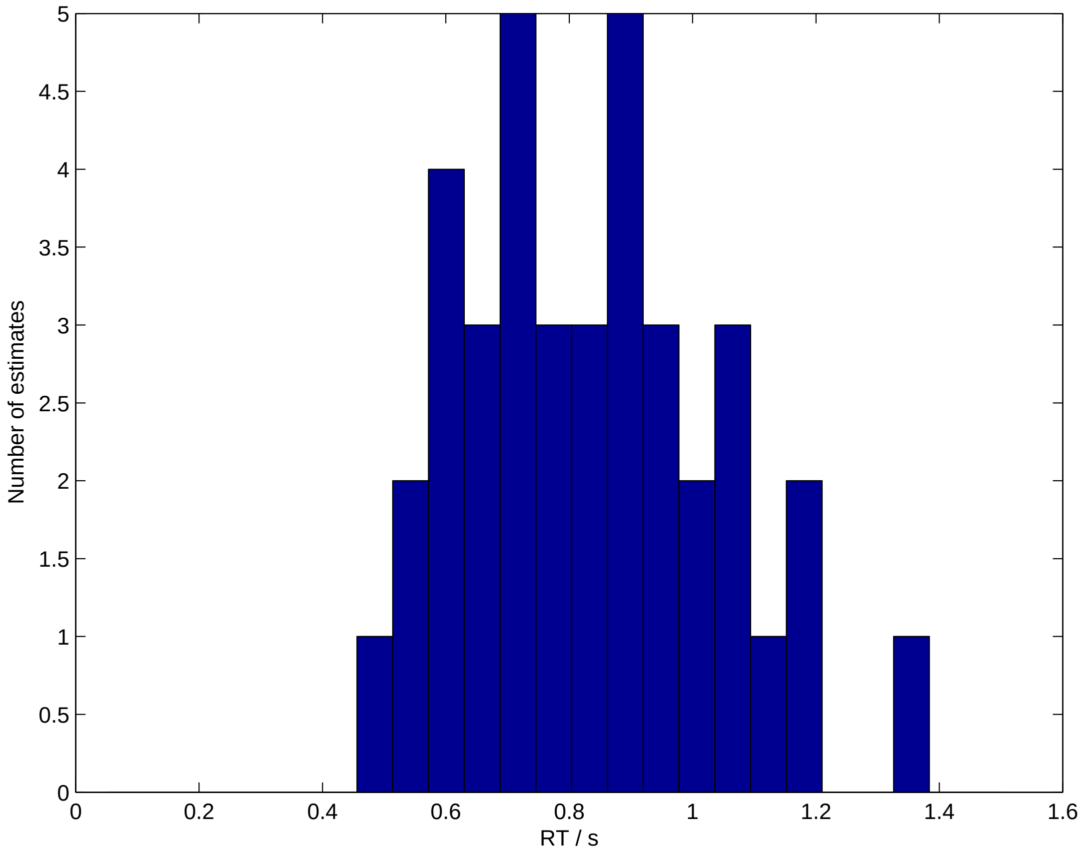

Figure 7: Histogram of T_{60} estimates for room A152, real recording, variable line fitting limits

Evaluation results

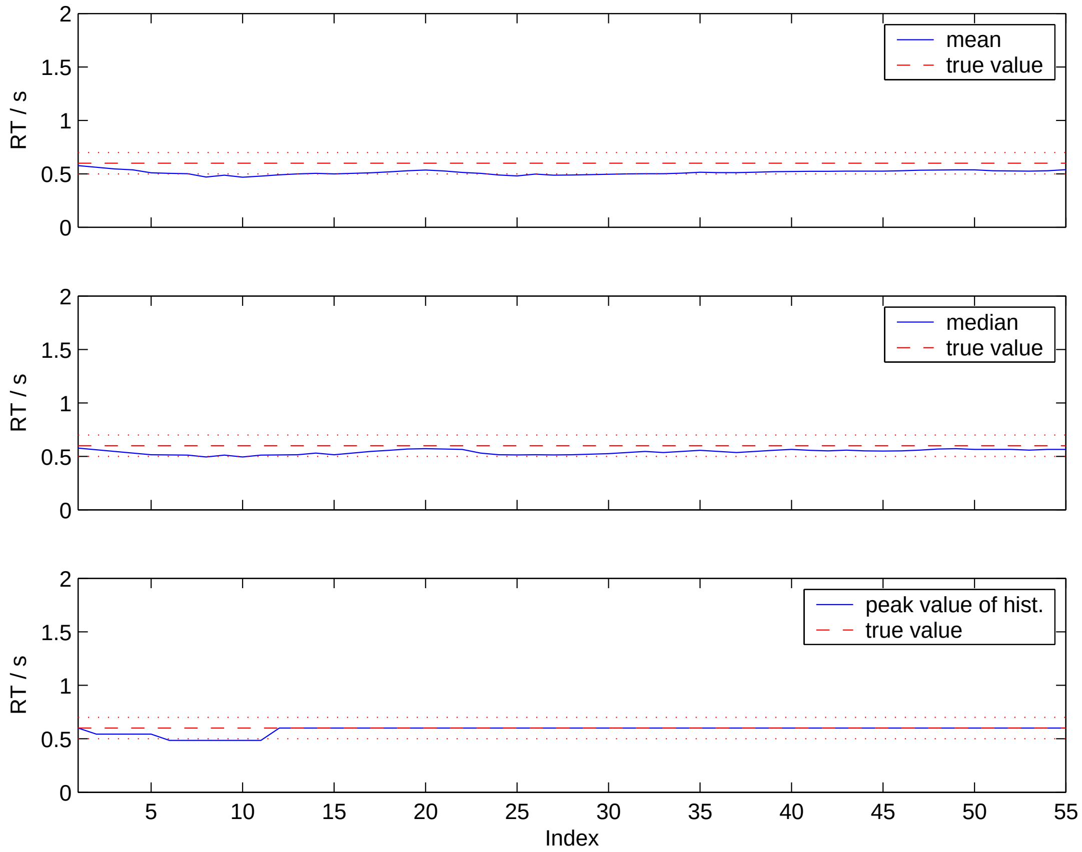

Figure 8: Three different statistics calculated from T_{60} estimates for room T3, real recording, variable line fitting limits

Evaluation results

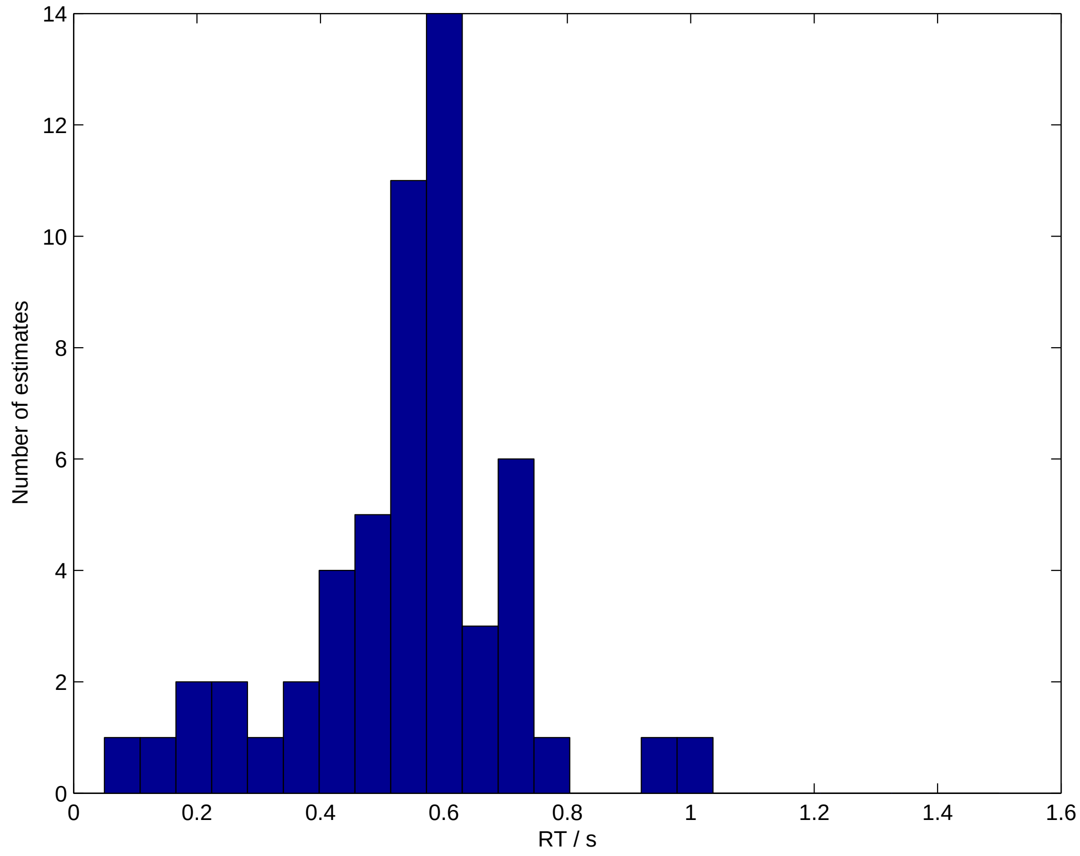

Figure 9: Histogram of T_{60} estimates for room T3, real recording, variable line fitting limits

Future work and conclusions

How to improve the algorithm performance?

- A clear downside is that the algorithm only works with sudden impulsive sounds $\longrightarrow$ improve the coarse segmentation part to detect all decaying segments with high enough SNR
- The algorithm is computationally quite heavy, some parts could possibly be left out
- The method performs well, matching human performance at its best

References

[1] Alexis Baskind and Olivier 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.

[2] W. T. Chu. Comparison of Reverberation Measurements Using Schroeder's Impulse Method and Decay-Curve Averaging Method. *Journal of The Acoustical Society of America*, 63(5):1444–1450, 1978.

[3] Laurent Couvreur, Christophe Ris, and Christophe Couvreur. Model-based Blind Estimation of Reverberation Time: Application

to Robust ASR in Reverberant Environments. In *Proceedings of the European Conference on Speech Communication and Technology (EUROSPEECH-2001)*, volume 1, pages 2631–2634, Aalborg, Denmark, September 2001.

[4] Trevor J. Cox and Francis F. Li nand Paul Darlington. Extracting Room Reverberation Time from Speech Using Artificial Neural Networks. *Journal of The Audio Engineering Society*, 49(4):219–230, April 2001.

[5] Martin Hansen. A Method for Calculating Reverberation Time from Musical Signals. Technical Report 60, The Acoustics Laboratory, Technical University of Denmark, Building 352, DK-2800 Lynbgy, 1995.

[6] Katia Lebart, Jean-Marc Boucher, and Philippe Denbigh. A New

Method Based on Spectral Subtraction for Speech Dereverberation. *Acustica/Acta Acustica*, 87(3):359–366, 2001.

[7] Anders Lundeby, Tor Erik Vigran, Heinrich Bietz, and Michael Vorländer. Uncertainties of Measurements in Room Acoustics. *Acustica*, 81:344–355, 1995. Dedicated to Prof. Dr. Heinrich Kuttruff on the occasion of his 65th birthday.

[8] Rama Ratnam, Douglas L. Jones, Bruce C. Wheeler, William D. O'Brien Jr., Charissa R. Lansing, and Albert S. Feng. Blind Estimation of Reverberation Time. *Journal of The Acoustical Society of America*, 114(5):2877–2892, November 2003.

[9] José Vieira. Automatic Estimation of Reverberation Time. In *Proceedings of the AES 116th International Convention*, Berlin, Germany, May 2004.

Bibliography

[10] Thomas Wittkopp. *Two-Channel Noise Reduction Algorithms Motivated by Models of Binaural Interaction*. PhD thesis, Carl von Ossietzky University Oldenburg, March 2001. http://docserver.bis.uni-oldenburg.de/publikationen/dissertation/2001/wittwo01/pdf/wittwo01.pdf.