Proceedings of 20th International Congress on Acoustics, ICA 2010

23–27 August 2010, Sydney, Australia

Blind Reverberation Time Estimation

Roman Scharrer and Michael Vorländer

Institute of Technical Acoustics, RWTH Aachen University, Germany

PACS: 43.60.Jn, 43.60.Np

ABSTRACT

This paper presents a method, which is able to give a blind estimation of the reverberation time of an enclosed space, using only the signal from two or more spatially distributed receivers, for example a binaural signal. There is no need for a controlled or known excitation signal, and there are no special requirements for the excitation signal. The method works with any kind of acoustic source, such as a speaking person, a musical instrument or a noise source. The indicator used for the reverberation time estimation is the spatial coherence and the coherence's dependency on the block size used for the coherence calculation. Using a neural network as estimator, a unique dependency between the block size dependent spatial coherence and the reverberation time could be verified and used for reverberation time estimation.

INTRODUCTION

State-of-the-art hearing aids, and other audio processing instruments, implement signal processing strategies tailored to the specific listening environments. These instruments are expected to have the ability to evaluate the characteristics of the environment, and accordingly use the most appropriate signal processing strategy [1]. Accordingly, a robust and reliable method to estimate the reverberation time from passively received microphone signals represents an important technology to improve the device's performance and the user experience.

Measurements and estimations of reverberation times have been the subject of many studies over the years. The reverberation time is an important and commonly quoted objective acoustic parameter for rooms. Reverberation influences speech intelligibility as well as music enjoyment. Reverberation also has a big influence on signal processing strategies, like beam forming, time delay estimation or noise suppression. Therefore, knowledge about the reverberation time can improve the quality of the results of such signal processing. The main problem in this context is the determination of the reverberation time with given limits, like uncontrolled excitation and an unknown acoustic environment. In a lot of situations, no controlled excitation is possible. This starts with occupied rooms, where people in the room would perceive the measurement sound, typically a sweep or noise, as annoying. Furthermore, for a lot of applications, no active excitation is possible at all. For example in transportable devices that do not include a speaker. In those cases only a blind reverberation time estimation, with no knowledge of the excitation signal itself, is possible. Especially the blind estimation of reverberation times is still a field with lots of uncertainties and room for improvement. Most methods only work for special conditions, as they often make certain assumptions on the unknown excitation signal or the room.

REVERBERATION TIME ESTIMATION

Measurements of reverberation time usually work with switched-off noise [16] or impulse response measurements [17]. The measurement procedure is standardized in [5], but in a lot of situations, no controlled excitation is possible.

Most methods for reverberation time estimation try to emulate the method of switched-off noise. In this case, a noise source excites a steady sound field in a room. After the noise source is switched off, the sound level in the room will decay linearly. An evaluation of this decay reveals the reverberation time [10]. The only difference for (semi blind) reverberation time estimation is that there is no control over the sound source. Some methods scan the audio signal for gaps and the level decay is evaluated [19]. Other methods are maximum likelihood estimation[15][23], neural networks [3] or blind source separation [21].

Blind source separation approaches use the effect, that for a correct source separation, the room impulse response is a by-product, which can directly be evaluated, for example using the methods described in [5]. However, this method has a critical drawback: deconvolution only works when the room impulse response is minimum phase, a condition that is not meet in most cases [15]. Therefore the method will not work in most environments.

Maximum likelihood methods usually try to estimate the reverberation time using the decay of the envelope of the autocorrelation function. Most of these methods have problems dealing with noise [12], or coupled rooms, where the level decay shows multiple decay rates [7].

A major drawback is that those methods can be fooled by using an excitation signal with reverberation. This leads to a signal showing two decay rates, that of the room as well as that of the reverberated signal itself. Accordingly, most methods will return bad estimations. Additionally, the methods that can deal with multiple decay rates will return the combined reverberation time of signal and room, where e.g. for dereverberation methods, the reverberation time of the room alone is critical and the reverberation in the source signal is unimportant and maybe even wanted.

There are also approaches using neural networks, which are trained with input signals for different rooms, though this method is not really blind as the network has to be trained with a known sound sequence. Reverberation time can only be estimated at occurrences of this sequence [3].

SPATIAL COHERENCE

Coherence in general is an indicator for the linearity and time-invariance of a system. The coherence of two signals x and y is the squared modulus of the cross spectral density normalized with the auto spectral densities:

$$\gamma_{xy}^2(\omega) = \frac{|S_{xy}(\omega)|^2}{S_{xx}(\omega)S_{yy}(\omega)} \tag{1}$$

For time discrete systems, the power spectral densities are usually estimated using a method like the 'Welch Periodogram' [22], which also includes the time averaging necessary for a valid coherence calculation [6]. This method segments the time signal into periods of a certain block size. For every segment the spectrum is calculated and averaged for the final spectral density estimation. The block size influences the duration of the single segments as well as the spectral resolution of the final power density estimation.

In room acoustics, spatial coherence is often used to validate the diffuseness of a sound field. The magnitude squared coherence between two pressure receivers with a distance r can be described as:

$$\gamma_{pp}^2(\omega,r) = \left(\frac{\sin(\omega r/c)}{\omega r/c}\right)^2 \tag{2}$$

These two assumptions are slightly flawed, as rooms are usually considered linear and time-invariant. A diffuse sound field, however, is often used to describe the reverberation or steady state sound field in a room. An ideal diffuse sound field is no linear and time invariant system which can be proven with a theoretic coherence which is not unity [6]. In real rooms, the theoretic prediction of the spatial coherence only concurs with measured coherences when following some rules. Spatial and temporal averaging is necessary and the spectral resolution of the energy density estimations must be smaller than $1/T$ where T is the reverberation time of the room. When the spectral resolution is bigger than $1/T$, which corresponds in time domain to a block size that is longer than the impulse response, coherence becomes unity as expected for LTI-systems [6].

This behavior already indicates a dependency of the calculated coherence of the used spectral resolution respectively block size. Figure 1 shows the binaural coherence in a room with a volume of 240 m^3, a reverberation time of 5 s at 1 kHz and a source to receiver distance of 3 m. For better visibility, the coherence is slightly smoothed. The block size dependency of the coherence is very obvious. For small block sizes the coherence behaves similar to the estimation of a diffuse sound field, whereas it increases with block size and finally becomes close to unity for block sizes longer than the impulse response. Additionally, the coherence increases for frequencies above 2 kHz as the reverberation time of the room decreases.

Figure 2 shows the mean binaural coherence in four different rooms as a function of PSD calculation block size. For all four rooms, the mean coherence is very close to unity for block sizes of 2^{20}, as is to be expected for linear time invariant systems. For smaller block sizes the mean coherence decreases.

It is obvious, that this function differs for the different rooms. Accordingly, assuming that this behavior is unique, it should be possible to determine the reverberation time from the binaural coherences for different block sizes. In the following sections, all values such as reverberation time or coherence will be used as mean values over a frequency range of 100 Hz to 10 kHz. This approach is taken to improve the perceptibility. There is no change in the method for frequency dependent estimations.

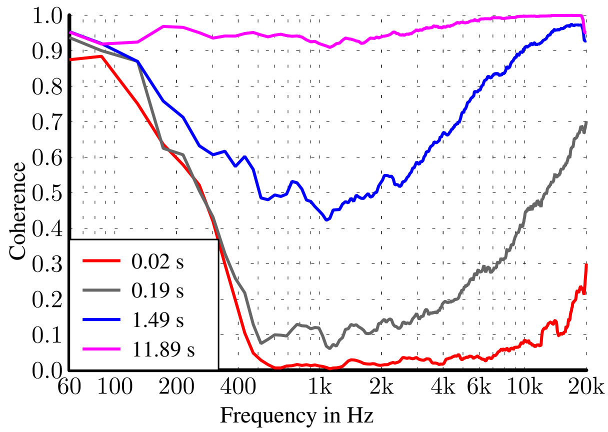

Figure 1: Binaural coherence in a reverberation room, calculated with four different time constants (block sizes) for power spectral density estimation

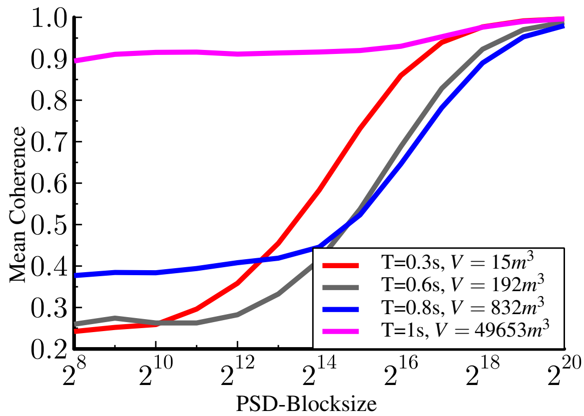

Figure 2: Mean binaural coherence in four different rooms as a function of PSD block size

We expect that some features of the room such as volume and reverberation time are a specific determinant of the shape of the spatial coherence curve. With the help of a neural network we will investigate how these quantities are hidden.

NEURAL NETWORKS

Artificial neural networks are designed as an abstraction of the signal processing in the human brain. A neural network consists of connected neurons. Neural networks are often used for pattern recognition and function interpolation. Usually they are designed as adaptive systems that change their structure based on information that is presented during a learning phase. They are very suitable to model complex or non-linear systems, with the advantage, that the complex model dependencies may be unknown. Additionally, neural networks are insensitive to single errors in the input data, making the process itself robust [8].

ESTIMATOR LAYOUT

The indicators used for the reverberation time estimation are the spatial coherences derived from power spectrums with different block or discrete-fourier-transformation (DFT) sizes.

The process of generating the input data for the estimation process is shown in figure 3. The coherence is calculated for

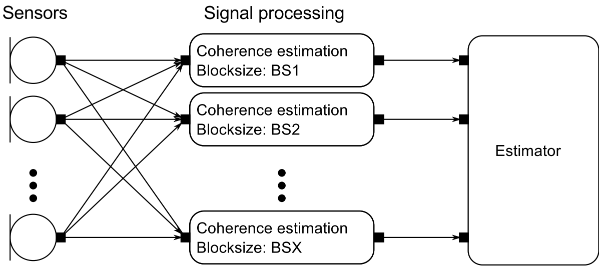

Figure 3: Sceme of the pre-processing. The coherence is calculated for all combinations of sensors and for different block sizes. The results are fed into the reverberation time estimator, in this case, a neural network

different combinations of input sensors and for different DFT block sizes. In this case binaural signals were used, so that there are only two input signals, one for each ear. The block sizes used are the powers of two in the range between 2^8 samples and 2^{20} samples. The signals used had a sampling rate of 44.1 kHz, so that the block sizes concur with time constants in the range of 5.8 ms to 23.8 s. The output of the coherence calculations are 13 frequency dependent values, similar to figure 1. At this stage, only broad band mean values are considered and the coherences are averaged throughout the interesting frequency range. This results in a total of 13 input values for the estimation process. Examples of these input values are shown in figure 2.

The results of those calculations are fed into the estimation process itself. One possible solution for the estimator is a neural network. A simple feed-forward network was chosen, including two hidden layers with 20 and 10 neurons per layer.

For frequency dependent estimations, there are two possibilities, either one network for every frequency of interest, or one network with one input for every frequency and block size. The second method has the advantage, that typical dependencies between the frequencies could be considered by the neural network.

TRAINING DATA

The network has to be trained once. This can be done with a set of measured or simulated input data and the according training targets. The training process has to be done for every microphone distribution. A simple way to generate sufficient amounts of training data are Monte-Carlo simulations. Monte-Carlo simulations are a method of repeated random calculations to understand the connection of input and output parameters of a modeled problem. They are especially helpful when the problem includes a great number of coupled degrees of freedom. In acoustics, Monte-Carlo simulations are often used to determine measurement uncertainties [14].

To limit the number of necessary simulations, the boundary conditions of the Monte-Carlo simulations have been chosen as follows: the room dimensions are chosen randomly in the range between 1 m and 200 m; the reverberation time is chosen randomly using a normal distribution with a mean value according to *DIN 18041* [4]. [4] describes proposed reverberation times for rooms with a given volume and purpose. It is meant as a guide in the acoustic layout of rooms for performances, education and other purposes. As a result, there is a slight correlation between reverberation time and room volume, as is to be expected for real rooms [9]. Nevertheless, there can be rooms with the same volume and different reverberation times and vice versa. The distance between source and receiver is selected randomly, limited by the room geometry. The number of sources is limited to one. The source signal is chosen randomly from a pool of signals, including noise, which is created individually each time, and a set of sound files, including speech as well as music. Using this random description, the impulse responses can be calculated with the stochastic room acoustic model described below. Afterwards the final receiver signal can be calculated by convolving the source signal with the impulse response and afterwards adding incoherent noise if a non-perfect signal to noise ratio is desired.

STOCHASTIC IMPULSE RESPONSES

Whenever a room acoustic measurement is not feasible or advisable, for instance due to time constraints or complexity, or because the room does not even exist, a room acoustic simulation is a possible solution for auralization or evaluation.

To determine the transfer function, or impulse response, from an acoustic source to a receiver in an enclosed space, a room acoustic model is necessary. There are vast differences between different methods, in terms of calculation time, audible accuracy and physical exactness. The applicability of some methods also depends on the room and the frequency range of interest. In general, wave based approaches are better suited for low frequency applications, with $f < f_c$

$$f_c = 2000\sqrt{\frac{T}{V}} \tag{3}$$

where f_c is the 'Schroeder Frequency', T is the reverberation time and V the volume of the room [18]. For higher frequencies with $f > f_c$, geometrical methods are better suited. There are also approaches of a combination of both methods to get realistic broadband impulse responses [2].

Most room acoustic simulation methods need a room model to work with. That means the room geometry as well as the boundary conditions have to be known. Another possibility is the stochastic simulation of a room. In this case no room model is necessary. The necessary information on the room are the room volume V, the room surface S, the source position, relative to the receiver and the average absorption coefficient α, the reflection factor R or the reverberation time of the room T. Furthermore, a directivity of the receiver could be specified, like a microphone directivity or a head related transfer function. The result is an impulse response as it could exists for such a room.

The stochastic room acoustic simulation is no approach to model an exact room but to find room impulse responses as they could exist in typical rooms. Using the same boundary conditions, a stochastic room acoustic simulation will return different impulse responses for every evaluation. This makes it rather useful for Monte-Carlo simulations. The method of stochastic room acoustic simulation is similar to the method of mirror sources. Therefore it is better suited for frequencies above the 'Schroeder Frequency', though due to time-frequency connections, it will also yield a realistic transfer function with discrete modes for lower frequencies [18].

The method of stochastic impulse responses is mostly based on a geometrical acoustics approach, as explained in [10] and [20].

The average temporal density of reflections arriving at time t in an arbitrary shaped room of the volume V can be expressed as:

$$\frac{\mathrm{d}N_\mathrm{r}}{\mathrm{d}t} = 4\pi\frac{c^3t^2}{V} \tag{4}$$

This means the density of sound reflections increases according to a quadratic law.

The total number of reflections in an impulse response of the length t_{ir} is

$$N_{\mathrm{ir}} = \int_0^{t_{\mathrm{ir}}} 4\pi \frac{c^3 t^2}{V} \mathrm{d}t \quad (5)$$

$$= \frac{4\pi}{3} \frac{c^3 t_{\mathrm{ir}}^3}{V} \quad (6)$$

Those reflections should be distributed in the interval $]t_{\mathrm{direct}} \;\; t_{\mathrm{ir}}]$ where t_{direct} is the time offset after which the direct sound reaches the receiver and prior to which no reflections can occur. A simple way of creating a random value with a quadratic distribution in the range $[0 \;\; t_{\mathrm{ir}}]$ is taking the cube root of an equal distribution P_{equal}, with $P_{\mathrm{equal}} \in [0 \;\; 1]$ and multiplying it with t_{ir}.

$$t_{\mathrm{ps}} = \sqrt[3]{P_{\mathrm{equal}}} \cdot t_{\mathrm{ir}} \quad (7)$$

The temporal reflection distribution can be transferred to a distribution of image source distances.

$$d_{\mathrm{ps}} = c \cdot t_{\mathrm{ps}} \quad (8)$$

The direction of the image sources relative to the head can be chosen randomly with an equal distribution on the full sphere.

An equal distribution on a sphere can be generated by the following method:

$$z \in [-1 \;\; 1] \quad (9)$$

$$t \in [0 \;\; 2\pi] \quad (10)$$

$$r = \sqrt{1 - z^2} \quad (11)$$

$$x = r \cdot \cos(t) \quad (12)$$

$$y = r \cdot \sin(t) \quad (13)$$

With the combination of random distances and random directions, a set of image sources can be generated. The transfer function from image source i to the receiver with the directivity H_{receiver} can be calculated as:

$$H_{\mathrm{i}}(\omega) = \frac{1}{ct_{\mathrm{i}}} \cdot H_{\mathrm{receiver}}(\omega, \phi, \theta) \cdot e^{\left[-j\omega - \frac{m(\omega)}{2} c + \overline{n} \ln(R(\omega))\right] t_{\mathrm{i}}} \quad (14)$$

$$\overline{n} = \frac{cS}{4V} \quad (15)$$

with t_{i} being the delay between the source and the receiver, m the air attenuation, $\overline{n}$ the average number of reflections per second for one sound ray and R the effective reflection factor. The values m and R will be frequency dependent for most situations.

The final transfer function is generated by a summation over all reflections and the direct sound which is calculated by the same formulas with $n = 0$.

$$H = H_{\mathrm{direct}} + \sum_{i=1}^{N_{\mathrm{ir}}} H_{\mathrm{i}} \quad (16)$$

RESULTS

Using the method of Monte-Carlo simulations, a set of 4000 different binaural signals was created, using the previously described boundary conditions. These signals were used to train and evaluate the neural network. Of the 4000 different rooms, 2000 were used for training, 1000 for verification of training results and 1000 for a final test. For every situation the binaural coherences for all block sizes were pre-calculated and saved for fast access, so that is was not necessary to calculate the coherences for every training step.

The training was applied using a Levenberg-Marquardt back-propagation approach [11][13]. This means that the training process was supervised and repeated until the gradient of the network performance undercut a certain threshold, meaning that only small improvements were to be gained by further training. The network performance is shown in figure 4. For every training epoch the relative mean squared error for the set of training, evaluation and test samples is shown. After 21 training epochs the minimal error in the validation and test samples is reached. The final relative squared error for the training samples is about 10^{-5}, the one for the validation and test samples about 10^{-3}.

Figure 5 shows the results from the training process, figure 6 the results of the verification and figure 7 the results of the final test. In each figure, every training sample is marked as a dot, indicating the neural networks output versus the real reverberation time which is the training target. The samples used for the final test have not been included in any training process, so they are new to the neural network whereas the other samples have all been presented in each training epoch. For all three training stages there is a very high correlation between neural network output and training target. The solid line indicates a linear least square error fit of the training results, the dotted line indicates the ideal result; training result equal to training target.

The neural network adapts very well to the presented training data and is able to model the relation between the coherence for different block sizes and the reverberation time. The relative error in all three sets is below 1 %, indicating a unique relation between the coherences for different block sizes and reverberation time.

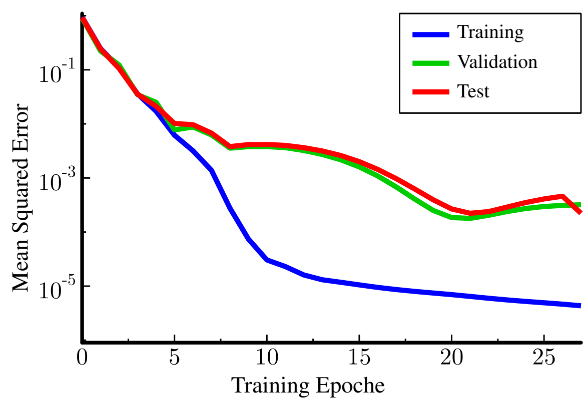

Figure 4: Relative mean squared error for the training epochs, separated into training, validation and test sets

Using this trained neural network, a test with recorded binaural signals returned the result shown in table 1. The rooms under test were two office rooms, a small and a big seminar room, a hallway and a reverberation chamber which had some absorbing foam on the floor to reduce the reverberation time. For the measurements, a dummy head was placed in the room, with an ear height of about 1.6 m. A loudspeaker was placed at a random distance of 2 to 4 m to the head, as the location allowed. Neither the dummy head nor the speaker were located close to a wall or other big objects. For the recording, the speaker played

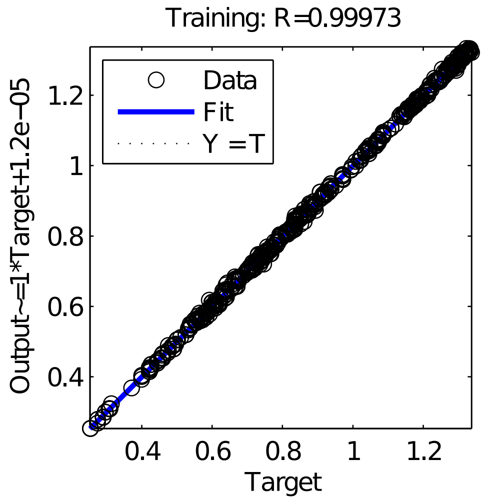

Figure 5: Results after the training of the neural network. The neural networks output is shown versus the training target. Each dot indicates a training sample. Additionally, a linear least squared error fit and the ideal fit are indicated.

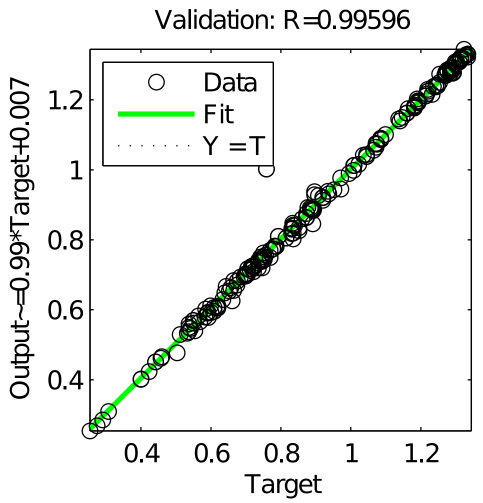

Figure 6: Results after the validation of the neural network. The neural networks output is shown versus the training target. Each dot indicates a training sample. Additionally, a linear least squared error fit and the ideal fit are indicated.

stochastic white noise which was recorded using the dummy head. This signal was then fed into the reverberation time estimation. For comparison, the impulse response of this specific setup was also measured for evaluation of the reverberation time according to [5].

There is a very good concordance between the estimation of the reverberation time and the measured reverberation time for the office rooms and the seminar rooms, though there are higher errors than in the training data. For the reverberation chamber, the estimation of the reverberation time is considerably too low. The reason is that the maximum reverberation time used for the training of the neural network was about 1.4 s. The estimation of about 1.7 s already exceeds this time, though not far enough. An extension of the training data should improve this estimation.

CONCLUSION

Using a neural network, the concept of the estimation of reverberation time from spatial coherence has proven to be possible.

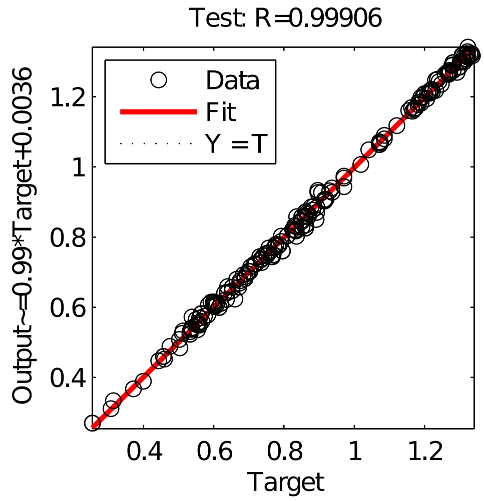

Figure 7: Results after a final test of the neural network. The neural networks output is shown versus the training target. Each dot indicates a training sample. Additionally, a linear least squared error fit and the ideal fit are indicated.

Table 1: Comparison of measured and estimated reverberation times

Room	T measured	T estimated
Office room 1	0.5 s	0.5 s
Office room 2	0.4 s	0.5 s
Small seminar room	0.6 s	0.7 s
Big seminar room	1.0 s	0.8 s
Hallway	1.4 s	1.3 s
Reverberation chamber	3.4 s	1.7 s

There is a unique dependency between the reverberation time and the block size dependent coherence between spatially distributed receivers. Using a neural network this dependency can be used to estimate the reverberation time of a room from e.g. a binaural signal. The network has to be trained once. To create this training data, a Monte-Carlo simulation using a stochastic room acoustic model has proven to be successful.

The training of the neural network showed a good adaption of the network within a limited number of training epochs. The relative error after the training was below 1 % for all training and test data. A simple feed forward network with two hidden layers was adequate for an estimation of broad band reverberation times from broad band coherences. For frequency dependent estimations, this setup has to be extended, with no change to the rest of the method.

A test with recorded signals shows a good concordance of the estimated reverberation times with measured reverberation times. This is limited by the amount of training data. A measurement in a reverberation room shows a poor performance of the estimator as the network has not been trained with such long reverberation times. An extension of the training data is necessary to include such situations as well.

The main difference of this method to existing solutions is, that the estimation parameters are estimated from spatial coherence instead of signal properties or signal shape. This leads to an independency from the excitation signals. There are no requirements to the excitation signal, like gaps which allow a slope decay evaluation. The method is also robust as neural networks are insensitive to errors in single input data and can be trained to return an estimation certainty. This can increase the robustness of the solution even further through post-processing of the result. One big advantage of this method is that the reverberation time is estimated based on properties of the sound field and not based on signal properties. That means, for example if a signal with a lot of reverberation is played back using a loudspeaker in a room with little reverberation, the reverberation time of the room is estimated whereas the reverberation in the signal itself is ignored.

The block size dependent spatial coherence also depends on other factors than the reverberation time, for example the signal to noise ratio. The next step is to examine if these relations are unique and whether or not it is also possible to estimate those values from the spatial coherence.

ACKNOWLEDGEMENT

The authors would like to thank Phonak for supporting this work.

REFERENCES

[1] S. Allegro, M. Büchler, and S. Launer. Automatic sound classification inspired by auditory scene analysis. In *Consistent and Reliable Acoustic Cues for Sound Analysis (CRAC)*, Aalborg, Denmark, 2001.

[2] M. Aretz, R. Nöthen, M. Vorländer, and D. Schröder. Combined Broadband Impulse Responses Using FEM and Hybrid Ray-Based Methods. In *EAA Auralization Symposium*, Helsinki, 2009.

[3] T. J. Cox, F. Li, and P. Darlington. Extracting room reverberation time from speech using artificial neural networks. *Journal of the Audio Engineering Society*, Vol. 49(4):219–230, 2001.

[4] DIN-18041. *Hörsamkeit in kleinen bis mittelgrossen Räumen*. Deutsches Institut für Normung e.V., Berlin, 2004.

[5] ISO-3382-2. *Acoustics - Measurement of room acoustic parameters - Part 2: Reverberation time in ordinary rooms*. International Organization for Standardization, 2008.

[6] F. Jacobsen and T. Roisin. The coherence of reverberant sound fields. *The Journal of the Acoustical Society of America*, Vol. 108(1):204–210, Juli 2000.

[7] P. Kendrick, F. F. Li, T. J. Cox, Y. Zhang, and J. A. Chambers. Blind estimation of reverberation parameters for non-diffuse rooms. *Acta Acustica United with Acustica*, Vol. 93(5):760–770, 2007.

[8] D. Kriesel. *A Brief Introduction to Neural Networks*. available at http://www.dkriesel.com, Mai 2007.

[9] M. Kuster. Reliability of estimating the room volume from a single room impulse response. *The Journal of the Acoustical Society of America*, Vol. 124(2):982–993, 2008.

[10] H. Kuttruff. *Room acoustics*. Elsevier, New York, 4th edition, 2000.

[11] K. Levenberg. A Method for the Solution of Certain Problems in Least Squares. *Quart. Appl. Math.*, Vol. 2:164–168, 1944.

[12] H. W. Löllmann and P. Vary. Estimation of the reverberation time in noisy environments. *Proc. Intl. Workshop Acoustic Echo and Noise Control*, 4, 2008.

[13] D. Marquardt. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM J. Appl. Math.*, Vol. 11:431–441, 1963.

[14] R. Martin, I. Witew, and M. Arana. Influence of the source orientation on the measurement of acoustic parameters. *Acta Acustica united Acoustica*, Vol. 93:387–397, Januar 2007.

[15] R. Ratnam, D. L. Jones, B. C. Wheeler, W. D. O. Jr, C. R. Lansing, and A. S. Feng. Blind estimation of reverberation time. *The Journal of the Acoustical Society of America*, Vol. 114(5):2877–2892, 2003.

[16] W. C. Sabine. Collected Papers on Acoustics. *Harvard U.P., Cambridge*, 1922.

[17] M. R. Schroeder. New Method of Measuring Reverberation Time. *The Journal of the Acoustical Society of America*, Vol. 37(6):1187–1188, 1965.

[18] M. R. Schroeder. The “Schroeder frequency” revisited. *The Journal of the Acoustical Society of America*, Vol. 99(5):3240–3241, Mai 1996.

[19] S. Vesa and A. Harma. Automatic estimation of reverberation time from binaural signals. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process (ICASSP)*, Philadelphia (Pennsylvania), USA, 2005.

[20] M. Vorländer. *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. Springer, Berlin, first edition, 2008.

[21] W. Wang, S. Sanei, and J. Chambers. Penalty function-based joint diagonalization approach for convolutive blind separation of nonstationary sources. *IEEE Transactions on Signal Processing*, Vol. 53(5):1654–1669, 2005.

[22] P. D. Welch. The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms. *IEEE Trans. Audio and Electroacoust.*, Vol. AU-15:70–73, 1967.

[23] Y. Zhang, J. A. Chambers, P. Kendrick, T. J. Cox, and F. F. Li. Blind estimation of reverberation time in occupied rooms. In *14th European Signal Processing Conference (EUSIPCO 2006)*, Florence, Italy, 2006.