Audible reflection density for different late reflection criteria in rooms

Krueger, Donata; Jeong, Cheol-Ho; Brunskog, Jonas; Buchholz, Jörg

Published in:
Internoise 2012

Publication date:
2012

Document Version
Publisher's PDF, also known as Version of record

Audible reflection density for different late reflection criteria in rooms

Donata Krueger[a)]
Cheol-Ho Jeong[b)]
Jonas Brunskog[c)]
Jörg Buchholz[d)]
Department of Electrical Engineering, Acoustic Technology, Technical University of Denmark
DK-2800 Kongens Lyngby, Denmark

For reasonably accurate but practical auralizations, some simplifications and approximations are needed. The main issue in the present investigation is that the reflection density of a room impulse response, in theory, increases so fast as a quadratic function of the elapsed time, even assuming only specular reflections. Therefore in this study, the upper threshold for audible reflection density is investigated for four different transition times of 25, 50, 75, and 100 ms through a headphone listening test. Binaural impulse responses and speech signals simulated in three rooms with different characteristics (an empty office, a lecture room, and an auditorium) are used as stimuli. Subjects are asked to increase/decrease the reflection density of a stimulus until they cannot distinguish it from the stimulus that follows the theoretical reflection density for the different transition times in the three rooms. When using binaural impulse responses, the upper limit of the audible reflection density turns out to be limited to 2800 reflections per second. For speech signals, the maximum audible reflection density is shown to be as low as 300 reflections per second, regardless of the room and transition time.

1 INTRODUCTION

In different fields of acoustics it is now commonplace to simulate an accurate sound field with its room acoustical parameters. The sound field can be described by a Room Impulse Response (RIR). It can be divided into three parts: direct sound, which is the sound that arrives straight from the source to the receiver; early reflections, which are the reflections that arrive

[a)] email: d.krueger@gta-akustik.de, current address: GTA mbH, Hannover, Germany
[b)] email: chj@elektro.dtu.dk
[c)] email: jbr@elektro.dtu.dk
[d)] email: Jorg.Buchholz@nal.gov.au, current address: National Acoustic Laboratories, Chatswood, Australia

early, generally speaking, within 50 or 80 ms from the direct sound and which are therefore important for a subjective evaluation of room acoustic qualities; and late reflections, which are the sound where numerous reflections overlap[1] and which is as well called diffuse exponential decaying reverberation tail[2]. In a universal physical sense, the number of reflections within a certain time interval, also called reflection density, increases proportionally to the square of the elapsed time, depending on the room volume[3]. After a certain time, e.g., the mixing time, the reflection density will be very large and the sound field will result into diffuse reverberation. By interfacing architectural acoustics to the human cognitive system, it is assumed that the auditory system reaches its temporal resolution at a certain number of reflections per second. The human ear follows the envelope of the signal and discriminations in the change and growth of the density cannot be done anymore. The time at which this happens is here called transition time.

This study presents the results of a master thesis study about only the late reflections content in the RIR. Psychoacoustic subjective tests were done to find a threshold where a change in the reflection density cannot be detected by the auditory system anymore. Hence, the parameters reflection density and transition time are regarded in an auditory sense and results are compared with physical assumptions.

2 THEORY

In geometrical room acoustics, sound propagation in a free field can graphically be drawn with a straight line as it can be assumed that sound propagates always in the fastest time respectively with the shortest way from sound source to sound receiver in a medium with equal propagation speed (similar to Fermat's principle known in the field of optics). If sound meets another medium, the free propagation is disturbed and a reflection occurs. Specular reflections follow the law that the angle of reflection is equal to the angle of incidence. The image source method is counting the number of reflections that will arrive in a certain time t if an impulse sound is emitted and can be described[4] by

$$N(t) = \frac{\frac{4}{3}\pi (ct)^3}{V}, \tag{1}$$

with room volume V in [m³] and volume of a sphere with radius ct, where c is the speed of sound [m/s] and t is the time [s]. Hence the tested parameter, reflection density, which is the number of reflections dN/dt within a time interval, grows proportionally to the square of the elapsed time by differentiating equation (1) to

$$\frac{dN}{dt} = 4\pi \frac{c^3}{V} t^2, \tag{2}$$

which is assumed to describe the temporal development of the reflection density for a measured RIR in a room. Polack[5] defined a mixing time as the time from when it is impossible to distinguish separate reflections in a pure physical sense[6]. The one used in this present investigation is also called mixing time which is further developed by Defrance. It is not just a measure in the pure physical sense. Perceptional information is included as the criterion takes the limited time resolution of the auditory system into account. An overlap of at least ten reflections within a characteristic time resolution of the auditory system, equal to 24 ms, is considered. It leads to following approximation

$$t_{mix,Defrance} \approx \sqrt{V}[ms]. \tag{3}$$

Combining equation (2) and (4) leads to

$$\frac{\Delta N}{\Delta t} = 4\pi c^3 \approx 500\frac{1}{s}. \tag{4}$$

Rearranging of equation (2) leads to

$$\Delta N = 4\pi \frac{c^3}{V} t^2 \Delta t. \tag{5}$$

Hence, assuming that there is on average only one reflection per time interval dt, so that in equation (4) $dN = 1$, time t_1 can be determined with

$$t_1 = \sqrt{\frac{V}{4\pi c^3 \Delta t}}. \tag{6}$$

3 METHODS

The goal of the study is to test the auditory perception of mixing time and reflection density in the late reverberation tail. Therefore, two RIRs are compared in a listening test. One signal is the reference signal which follows an increasing reflection density in the whole signal. The simulated signal is the test signal, which follows an increasing reflection density up to a certain transition time and a constant reflection density afterwards. Hence, the simulated signal is the same signal as the reference signal up to the transition time. Its constant reflection density value in the late tail is changed during the listening test depending on test subject responses until the same sound image is achieved by test and reference signal.

3.1 Real RIRs

The signals used in the listening test should be as close as possible to a real RIR. Three rooms are chosen with different room volume: An Office with a volume of 92.88 m³, a Lecture Room with a volume of 190.75 m³ and an Auditorium with a volume of 1174 m³. The RIR is measured by exciting the rooms with an exponential sweep of frequency range 20 Hz to 22.4 kHz, a stationary and broadband signal. Due to the frequency characteristics of the dodecahedron loudspeaker used for the measurements, RIRs are filtered to a frequency range of 100 Hz to 10 kHz by using the 6th order butterworth filter with mentioned frequencies as cutoffs.

Room acoustical parameters are determined by using the integrated impulse response method[7]. The reverberation time (RT), early decay time (EDT), clarity (C80) and definition (D50) are evaluated to get an objective impression of the rooms. In addition, the mixing time and increase of the theoretical reflection density is calculated as given by equations (2) and (4). Values for the Defrance’s mixing time are given in table 1.

3.2 Simulated signals

The Poisson process can be used as a statistical tool for the simulation of a reflection density. It is a stochastic process in which “events” (reflections) occur continuously and independently from each other. It can be described with $\{N(t): t \geq 0\}$ with $N(t)$ as the number of events that occur up to a time starting from 0 s. Properties considered using the Poisson process are that $N(0)=0$, the number of events in non-overlapping intervals are independent from each other, the number of events in a time interval has a Poisson distribution, more precisely: The probability of having exactly k events is given by

$$f(k;\lambda)=\frac{\lambda^k e^{-\lambda}}{k!}, \tag{7}$$

where $k!$ is the factorial of k, e is the base of the natural logarithm and λ is equal to the expected number of events that occur during a given interval. Therefore $\lambda = dN/dt$.

Examples for three different values for λ are given in figure 1. It needs to be pointed out that resulting event histories are independent of a special room. The Poisson process has only been implemented for constant values of the reflection density, which is a very good solution for creating the constant reflection density for the late reverberation tail of the test signal. A step-wise increasing reflection density is implemented in Matlab as an increasing reflection density is asked for the reference signal and the first part of the test signal. Exemplary resulting patterns are shown in figure 2. In figure 3 the created patterns are normalized towards the "zero-line" as the same is done for a real RIR. Therefore, the mean value of each time interval in the reflection density pattern is subtracted from it.

In a realistic environment reflections are not just pure peaks. The reflected sound contains the frequency content of the given impulse by taking absorption, diffusion and scattering of the boundary into account. In this study boundary conditions will not be regarded and each reflection contains the same power spectrum. For this reasons, the event history is convolved with the response of the dodecahedron, which is separately measured in an anechoic chamber. Hence, the simulated signals are forced to have the same direct sound than the measured RIRs in the three rooms, Office/Lecture/Auditorium and show a more realistic envelope characteristic than the reflection density pattern before.

However, the envelopes of the simulated signals are adjusted towards the measured RIRs ones to achieve more realistic energy decay. This is done by using a complex All-Pole ERB filter bank according to Hohmann[8], which divides the signal in different frequency channels and makes it possible to adjust the frequency content of the simulated to the measured RIR more precise. The used filter bank is based on the theory of the nonlinearity spaced auditory filter bank of the auditory system. It can be described by the concept of the equivalent rectangular bandwidth of the auditory filters of the cochlea. The resulting 44 filters are arranged to adjust phase, gain and time delay so that for a delta impulse input the summed real output approximates the input impulse. An approximation of the Hilbert envelope for each of the channels can be derived by taking the absolute value of the complex output of the filter. The so-called "correction envelope" is determined by dividing the envelope of the measured RIR by the envelope of the simulated RIR. By multiplying it to the real signal for each channel and summing up the 44 channels, the simulated RIR is created with corrected envelope.

The simulation of the reference signal with increasing reflection density is done. In a last step, the test signal with increasing reflection density up to a certain transition time and a constant reflection density afterwards needs to be assembled. Chosen transition times are 25 ms, 50 ms, 75 ms and 100 ms. Figure 5 shows an example of assembling. In figure 5a the reference signal with growing reflection density is shown for the Lecture room. This signal is windowed at a transition time of 50 ms, which is shown in figure 5b. The windowing is necessary to reduce "clicking noise" which could appear due to phase shifts in the assembled signal. The windows are shown in figure 6. In figure 5c the windowed signal with constant reflection density of 100 reflections per second is shown. The resulting test signal is achieved by assembling the signals from 5b and 5c. Hence, it follows the growing reflection density up to the transition time and a constant reflection density afterwards.

3.3 Psychoacoustics

The listening test is done with a 3-Alternative Forced Choice method in conjunction with the one-up two-down method. Therefore, the threshold is given by the point where the hit rate is 70.7 %. In addition to the RIRs, test subjects have to listen to configurations with SPEECH. The RIRs for the three rooms are combined with SPEECH, as listening to RIRs is not a usual sound for humans. For the SPEECH track, Danish sentences are picked randomly out of 180 possibilities, which are supposed to have the same long-term spectrum[9]. For each of the conditions RIR and SPEECH, the four transition times (0.025 s, 0.050 s, 0.075 s, 0.100 s) after which the reflection density is created with a constant value are tested. Each time is tested three times to control reliability of the reached threshold for each test subject.

Six non-danish, male university students are chosen to be test subjects. They are in the age between 23 and 26 and had normal hearing. The test was divided into several sessions so that the concentration is kept over testing time. The test subjects do the headphone listening test in a sound attenuated booth by using a diotic signal at 55 dB FS.

3.4 Statistics

The results of the listening tests are treated with several statistical methods[10]. By determine quartiles with a cumulative distribution function, outliers can be determined and neglected and the median value is taken for the threshold as it is more robust than a mean value towards outliers. A simple linear regression is used to estimate if there is a linear relationship between the value of the reflection density and the transition time and if these parameters depend linearly from each other. Therefore, the median over all test subjects is expressed as a linear function over the transition time. To get the best fitted regression line to the data, the principle of least squares is used. Residuals are determined. The goodness of fit is evaluated by the determination of an ANOVA (analysis of variance) table. Parameters are the regression parameters β_0 and β_1, the coefficient of determination R^2, an adjusted coefficient of determination R_a^2 and a statistical hypothesis F-test with null hypothesis which leads to a p-value. The R_a^2-values should be close to one, the p-values should be close to zero.

4 RESULTS

Figure 7 shows the median values over all test subjects for the three rooms for both conditions, RIR in a), SPEECH in b). The Office is indicated with a symbol ‘o’, the Lecture Room with a symbol ‘x’ and the Auditorium with a symbol ‘*’. In figure 7a it can clearly be seen that median values for the Office are higher than for Auditorium and Lecture Room. This is as well observable in figure 7b but the trend that values of the Auditorium are clearly higher than the ones of the Lecture Room cannot be made in this case.

Simple linear regression lines are added in the figures. The reflection density decreases over time for all conditions except testing RIR condition in the Auditorium. Here, an increase in the reflection density over time can be observed. Before going into deeper evaluation, the linear regression lines are evaluated towards the goodness of fit.

5 DISCUSSION

The goodness of fit is done by an ANOVA table; values are given in table 2. The outcome is that on the one hand the linear regression line is not a good fitting for the RIR, but on the other hand it is a good fitting for SPEECH.

Using a two-way ANOVA (see table 3), where the means of the needed reflection density over the transition times are compared to the means of the reflection density of the three rooms, shows that transitions time do not affect the needed reflection density as opposite to the rooms and their characteristics for the RIR case. No evidence of an interaction can be observed between the transition time and the rooms. For SPEECH, both transition times and the rooms affect the needed reflection density and again, there is no evidence of an interaction between transition times and rooms. This leads to the conclusion that the room characteristics affect the needed reflection density.

As the room characteristic is described by different room acoustical parameter a correlation matrix is formed to evaluate which room acoustical parameter does affect the reflection density most, shown in table 4. As the linear fitting is only acceptable for SPEECH, the correlation matrix is only done for this configuration. It is shown that the reverberation time has the biggest influence on the number of reflections so that the higher the reverberation is, the higher is the necessary reflection density.

In figure 8, test results are compared to the theoretical density curves given in figure 1. Theoretical values are much higher than auditory values, which seem to stagnate around a certain value. In comparison with some of the previous studies, the measured reflection density for RIR deviates with values below 1500 quiet strongly. It is even lower for SPEECH with values below 300 reflections per second.

However, a comparison to other authors cannot easily be done as they used different signal processing tools and room characteristics. Cremer[4] had suggested a mean density of 2800 reflections to achieve a sound impression to the human ear which is as good as simulating a signal following the physical growth of the reflection density. This value can be taken as an upper limit for the temporal resolution of the auditory system. The time value when a minimum reflection density of 2800 is achieved can be taken from the theoretical reflection density curve in figure 1. Values for the three rooms are given in table 5.

A comparison between these values and the ones in table 1 shows that Defrance's mixing time is nearly four times bigger than the transition time considering the temporal resolution of the auditory system. This means that for the Office the chosen transition times lie all above the t_{2800}-value of 22 ms. For the Lecture Room the t_{2800}-value of 32 ms lies slightly above the lowest transition time of 25 ms and for the Auditorium the t_{2800}-value of 81 ms lies above the tested transition time of 25 ms, 50 ms, 75 ms.

6 CONCLUSIONS AND FURTHER WORK

This study is concerned with the auditory perception of reflection density and examination of the mixing time. A simulation method has been developed so that the test signal is composed of the growing theoretical reflection density up to a certain transition time and a constant reflection density afterwards, which is created with the Poisson process, whereas the reference signal follows the theoretical reflection density. The envelope was adjusted towards the envelope of the measured RIRs of three rooms with different room characteristics (Lecture Room, Office, Auditorium). Chosen transition times were 25 ms, 50 ms, 75 ms and 100 ms. A listening test session was performed and the results were analyzed and discussed with following outcome:

For RIR: The reflection density is strongly correlated to the room volume. The temporal resolution of the auditory system limits the transition time towards a shorter time than the mixing time. A statistical description becomes possible around a reflection density of 2800 for auditory purpose.

For SPEECH: The reflection density decreases with increasing transition time. The reflection density takes values below 300 for all conditions. In general, the reflection density is almost ten times smaller for SPEECH than for RIR in comparison to the theoretical reflection density.

In the future, the listening test should be extended to transition times of 125 ms and 250 ms to observe a stagnating reflection density. The t_{2800}-value does not take the room characteristics into account. This can be a reason why the thresholds of the listening tests are clearly lower than 2800. According to the ANOVA tests, it is shown that the reflection density is not only depending on the room volume, it is as well linked to the reverberation time. Therefore, listening tests with more rooms with different reverberation times should be done to confirm the dependency of reflection density on reverberation time. In addition, the influence of strong early reflections on the reflection density should be tested. Beside RIR and SPEECH, the reflection density should be tested with MUSIC.

6 REFERENCES

1. S. Hameed, J. Pakarinen, K. Valde and V. Pulkki, ”Psychoacoustic Cues in Room Size Perception”, *Audio Engineering Society*, 116th Convention, Berlin, Germany, Convention Paper **6084** (2004).

2. P. Rubak and L.G. Johansen, ”Artificial Reverberation based on a Pseudo-random Impulse Response II, *Audio Engineering Society*, 106th Convention, Munich, Germany, Convention Paper **4900** (1999).

3. J. E. Summers, R. R. Torres, Y. Shimizu and B.-I. L. Dalenbäck, *”Adapting a randomized beamaxis-tracing algorithm to modeling of coupled rooms via late-part ray tracing”, J. Acoust. Soc. of Am.*, **188**(3), 1491-1502 (2005).

4. L. Cremer and H.A. Müller, *”Die wissenschaftlichen Grundlagen der Raumakustik – Band I”*, S. Hirzel Verlag, Stuttgart, Germany, 2nd Edition, ISBN: 3-7776-0315-5 (1978).

5. G. Defrance and J-D. Polack, ”Measuring the mixing time in auditoria“, Conference Paper, *Acoustic08*, Paris (2008).

6. C.-H. Jeong, J. Brunskog and F. Jacobsen, ”Room acoustic transition time based on reflection overlap (L)“, *J. Acoust. Soc.Am.*, **127**(5), 2733-2736 (2010).

7. A. Farina, ”Simultaneous measurement of impulse response and distortion with a swept-sine technique”, *108th AES Convention*, Paris, preprint **5093**(D-4) (1996).

8. V. Hohmann, ”Frequency analysis and synthesis using a Gammatone filterbank”, *Acta Acustica*, **88**, 433-422 (2002).

9. T. Dau and J.B. Nielsen, ”Development of a Danish speech intelligibility test“, *International Journal of Audiology*, in press at the time of printing (2009).

10. L. Papula, *”Mathematik für Ingenieure und Naturwissenschaftler – Band 3*, Vieweg, Germany, 4th Edition, ISBN: 3-528-34937-9 (2001).

Table 1 – t_{mix} for three chosen rooms (Office, Lecture and Auditorium) calculated with equation 4.

Room	t_{mix} [s]
Office	0.13
Lecture	0.09
Auditorium	0.32

Table 2 – Outcome of the Anova table.

	R_a^2, RIR	R_a^2, SPEECH	p-value, RIR	p-value, SPEECH
Lecture	-0.358	0.910	0.692	0.031
Office	-0.440	0.919	0.800	0.027
Auditorium	0.100	0.460	0.368	0.200

Table 3 – Outcome of the two-way Anova table considering reflection density over transition time and rooms.

	p-value, RIR	p-value, SPEECH
Columns	0.4313	0.0007
Rows	0	0
Interaction	0.7924	0.7884

Table 4 – Time value for minimum reflection density value of 2800.

Room	t_{2800} [s]
Office	0.032
Lecture	0.022
Auditorium	0.081

Table 5 – Correlation matrix between fitting parameters and room characteristics for SPEECH configuration, only sample correlations with a significance level up to 20% are shown.

	$\beta_{,0}$	$\beta_{,1}$	V	V_{log}	V_{cub}	Avg_{T20}	Avg_{T30}	Avg_{EDT}	Avg_{C80}	Avg_{D80}
$\beta_{,0}$	1.00	-	-	-	-	-	-	-	-	-
$\beta_{,1}$	-	1.00	-	-	-	0.99	0.99	0.98	-	-
V	-	-	1.00	0.98	0.99	-	-	-	-	-
V_{log}	-	-	-	1.00	0.96	-	-	-	-	-
V_{cub}	-	-	-	-	1.00	-	-	-	-	-

Avg_{T20}	-	-	-	-	-	1.00	1.00	0.99	-	-
Avg_{T30}	-	-	-	-	-	-	1.00	0.99	-	-
Avg_{EDT}	-	-	-	-	-	-	-	1.00	-	-
Avg_{C80}	-	-	-	-	-	-	-	-	1.00	0.97
Avg_{D50}	-	-	-	-	-	-	-	-	-	1.00

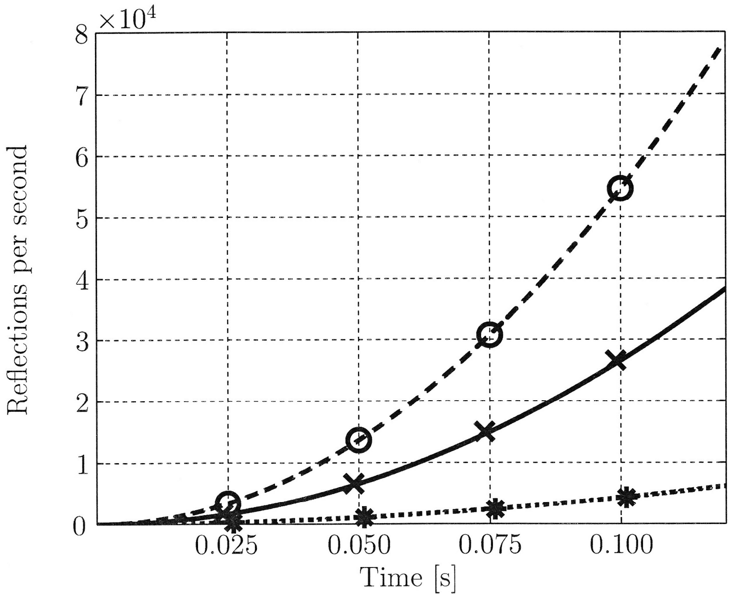

Fig. 1 – Theoretical reflection density curve for Office (dashed line, 'o'), Lecture Room (solid line, 'x') and Auditorium (dotted line, '').*

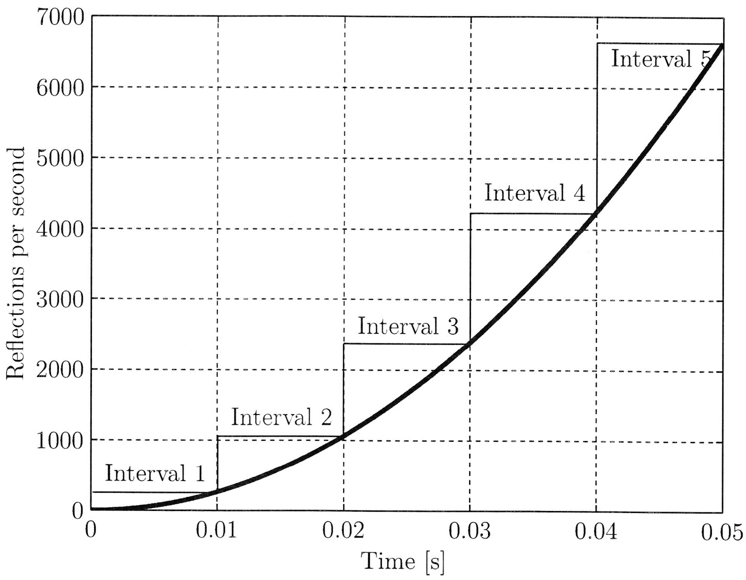

Fig. 2 – Theoretical reflection density curve for Lecture Room (solid line), curve is devided in intervals of 10 ms, value at each multiple of 10 ms-point is used as λ–values in poisson process

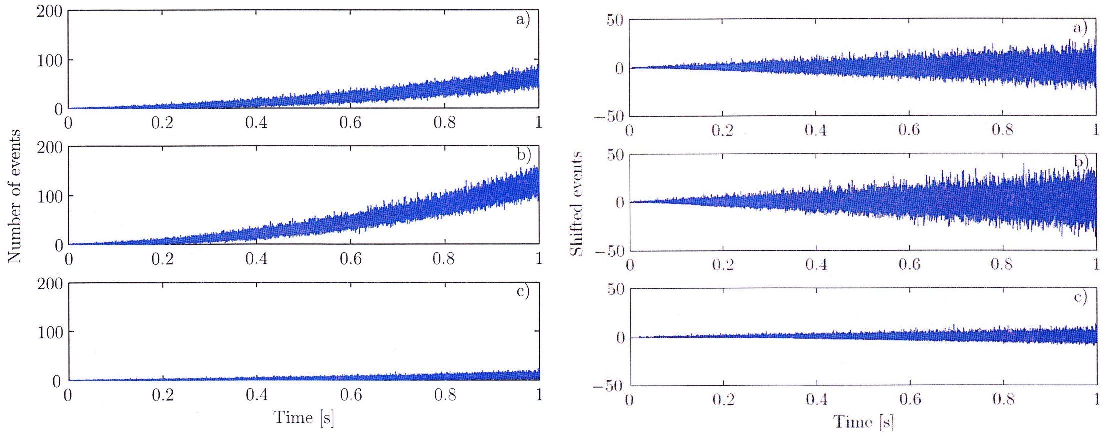

Fig. 3 – Event history when simulating an increasing reflection density before (left) and after (right) normalizing to the "zero-line", a) Lecture Room, b) Office, c) Auditorium

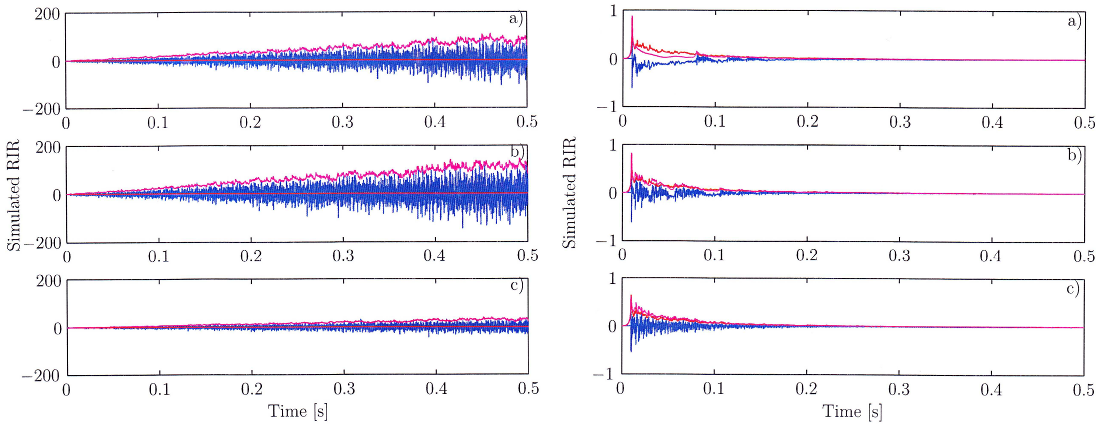

Fig. 4 – Before (left) and after (right) envelope adjustment: RIR with growing reflection density (blue line), Envelope of RIR with increasing reflection density (magenta line), Envelope of measured RIR of Lecture room (red line); a) Lecture room, b) Office, c) Auditorium.

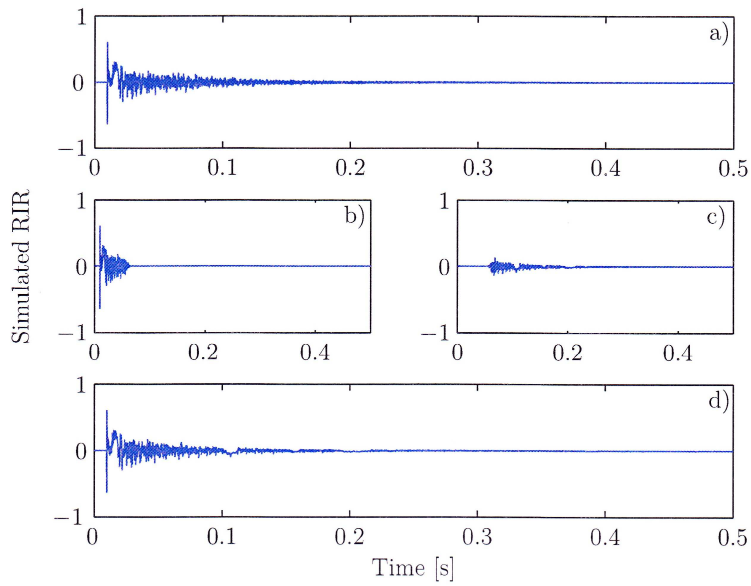

Fig. 5 – Example of a signal assembling at a transition time of 50 ms for the Lecture Room: a) Reference signal with growing reflection density, b) Windowed reference signal with growing reflection density, c) Windowed signal with constant reflection density, d) Test signal: assembling b)+c)

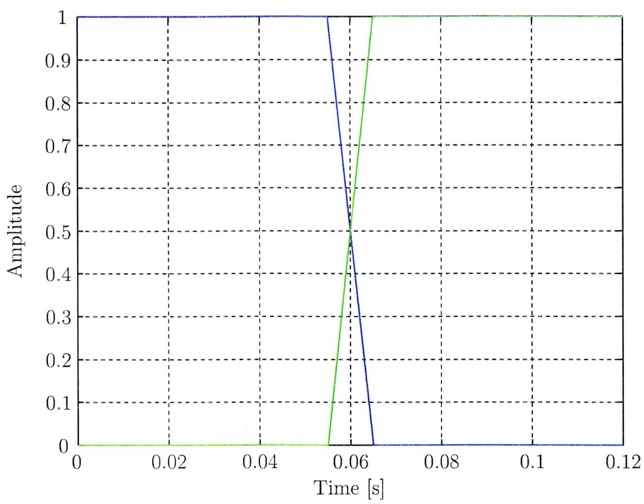

Fig. 6 – Example of windows for a transition time of 50 ms. Signals which should be windowed, have a time delay of 10 ms which is an outcome of the envelope adjustment. For this reason, the windows include a time delay of 10 ms and the crossing point at 50% is shifted to 60 ms. In total, the crossing time is kept at 10 ms for all transition times. The window marked with blue line is used for signal with increasing reflection density (figure 5b); the window marked with the green line is used for signal with constant reflection density (figure 5c).

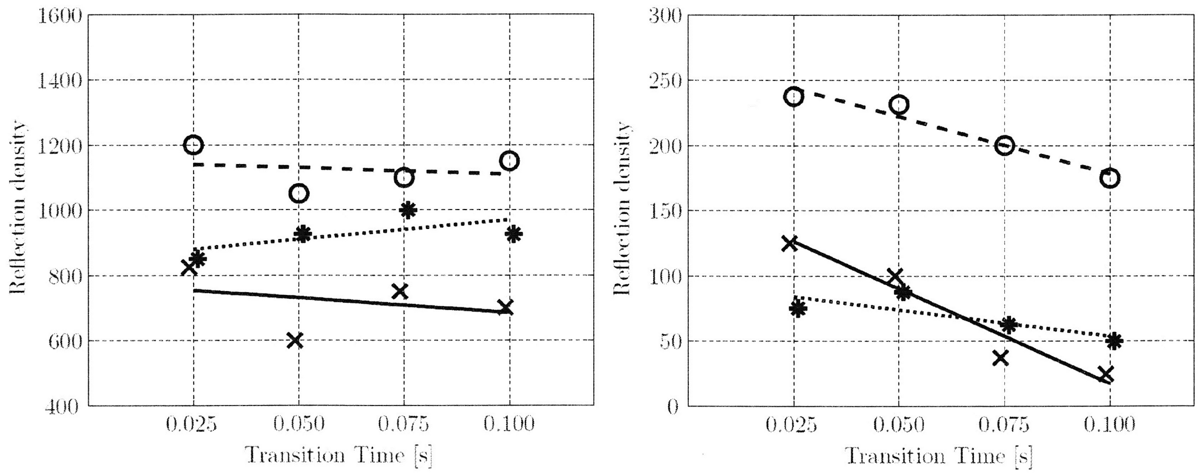

Fig. 7 – Simple linear regression for listening test with a) RIR and b) SPEECH (Lecture room 'x' and 'solid line', Office 'o' and 'dashed line', Auditorium '' and 'dotted line')*

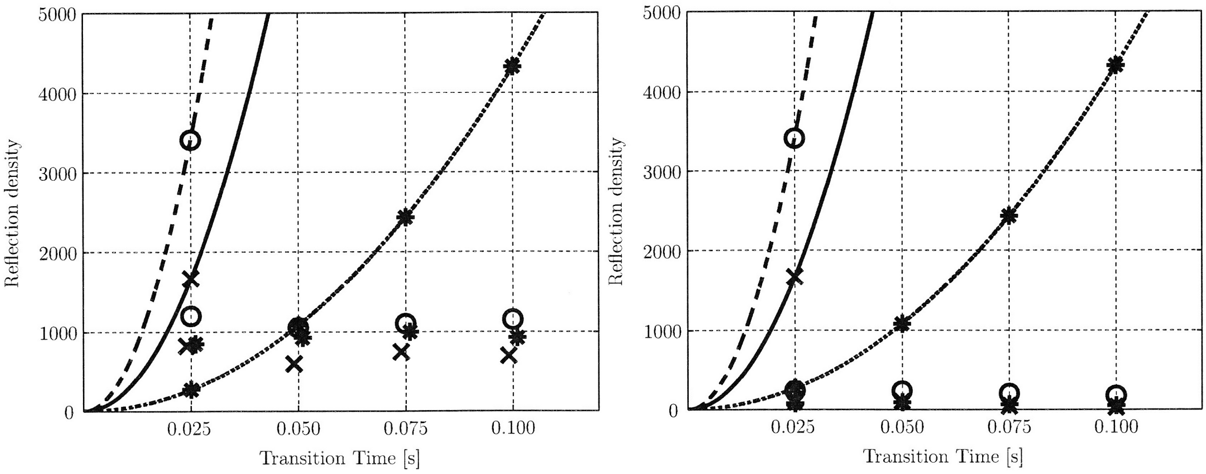

Fig. 8 – Comparison of theoretical density curve and measured threshold (Lecture room 'x' and 'solid line', Office 'o' and 'dashed line', Auditorium '' and 'dotted line').*