Analysis of room transfer function and reverberant signal statistics

Eleftheria Georganti[a], John Mourjopoulos[b] and Finn Jacobsen[a]

[a]Acoustic Technology Department, Technical University of Denmark, Ørsted Plads, B352, DK-2800 Lyngby, Denmark

[b]Audio and Acoustic Technology Group, Wire Communications Laboratory, Electrical Engineering & Computer Technology Department, University of Patras, 26500 Patras, Greece

ege@oersted.dtu.dk

This work examines existing statistical time-frequency models and techniques for Room Transfer Function (RTF) analysis (i.e. Schroeder's stochastic model and the standard deviation over frequency bands for the RTF magnitude). RTF fractional octave smoothing, as with 1/3 octave analysis, may lead to RTF simplifications that can be useful for audio applications, and this work examines the relationship of such operations with respect to the original RTF statistics. More specifically, the RTF statistics, derived after complex smoothing using 1/3 fractional octave analysis, are compared to the original statistics across frequency and across space inside typical rooms, by varying the source, the receiver position and the corresponding ratio of the direct and reverberant signal. In addition, this work examines the statistical quantities for speech and audio signals prior to their reproduction within rooms and when recorded in rooms. Histograms are used to compare RTF minima of typical "anechoic" and "reverberant" audio and speech signals, in order to model the alterations due to room acoustics.

1   Introduction

For many years now, statistical analysis has been a valuable tool in analyzing Room Transfer Functions (RTFs). Schroeder proposed a stochastic model [1] in the 1950s that was further developed by Polack [2, 3]. A good overview of the above models is given in [4]. Statistical models can lead to RTF simplifications that have been proved to be useful for several audio applications, such as room compensation, room modeling, auralisation and dereverberation techniques [4, 5]. The aim of this work is to examine statistical properties of measured RTFs across frequency and space inside typical rooms, when varying the distance between the position of source and receiver and the corresponding ratio of the direct and reverberant energy. More specifically, the standard deviation (in dB) over 1/3 octave frequency bands for the RTF magnitude is calculated and presented as a function of frequency and space.

The necessity of having a sufficient bandwidth in order to obtain values close to the expected value of 5.6 dB is shown (5.57 dB corresponding to the value of the standard deviation derived by Schroeder [1], when RTF normal modes overlap in frequency and there is negligible coherent transmission). When using higher fractional octave analysis (e.g. 1/16 or 1/64) the values for the standard deviation are shown to decrease.

An alternative way to obtain simplified versions of RTFs, via complex smoothing [6, 7] using 1/3 octave analysis, is investigated. The standard deviation over fractional octave bands is examined also for complex smoothed versions of the above measured RTFs.

This work also examines the statistical properties and distributions, using histograms to compare RTF minima of typical anechoic and reverberant audio-speech signals. Histograms are used to plot the RTF magnitude statistics for different distances between the source and the receiver in a room and the corresponding statistics of reverberant signals (when reproduced at the same points of the room). Furthermore the statistics of different types of anechoic signals (orchestral music, speech) are examined and compared to the corresponding reverberant signals.

2   Theory

2.1   Frequency domain statistical model

For intervals in the room impulse response when the echo density becomes high and at frequencies having high modal overlap, the statistical models of RTF as developed by Schroeder [1] in the frequency domain and more recently by Polack [2] in the time domain may be employed.

This implies that at such higher frequencies, the normal modes of a room overlap, and any source signal will simultaneously excite several room modes. Assuming a sine wave excitation and a microphone located in the reverberant field, the signal captured is the sum of contributions of large number of modes, where the real and the imaginary parts of the complex sound pressure can be considered as independent Gaussian processes that have the same variance [1]. This two-dimensional Gaussian density arises from the central limit theorem, assuming independence between the modes and implies that the magnitude frequency response follows a Rayleigh distribution.

These statistical properties are valid irrespective of the microphone position (provided that the direct sound is negligible compared to the reflected one) and irrespective of the room dimensions and properties at frequencies above Schroeder's frequency [1, 8],

$$f_{Schroeder} \approx 2000\sqrt{\frac{RT}{V}} \quad \text{(Hz)} \qquad (1)$$

where RT (sec) is the reverberation time and V (m³) is the volume of the room.

Additionally, Schroeder [1] has shown that at distances far from the sound source the standard deviation of the sound pressure level with respect to frequency is 5.57 dB. Diestel [9] has shown that the probability distribution of sound pressure is related to the ratio of the direct to reverberant sound. J. J. Jetzt [10] derived a function that relates the standard deviation of the magnitude of the frequency response to the ratio of the direct and reverberant energy, and he proposed the "standard deviation method" for measuring the critical distance.

2.2   Time domain statistical model

It is known that room impulse response tail may be simulated by exponentially decaying white noise [4]. Based on this, Polack developed a time-domain model complementing Schroeder's frequency-domain model, where he described the room impulse response tail as a realization of a non-stationary stochastic process:

$$h(t) = b(t)e^{-\delta t}, \quad t \geq 0 \qquad (2)$$

where b(t) is centered stationary Gaussian noise, and δ is:

$$\delta = \frac{3\ln 10}{RT} \qquad (3)$$

As Schroeder's model is valid above Schroeder's frequency (equation (1)), Polack's model is valid after a specific time interval from the initiation of the impulse response. This is so, because the time-domain response can only be Gaussian if a sufficient number of reflections overlap at any time instance. Since the reflection density increases, a value $t_{mixing} \approx \sqrt{V}$ (ms) was proposed [2] as the transition time between early reflections and late reverberation.

The previously described statistical models have introduced a useful framework for analysis and modeling of the room responses (RTFs) that is the main concern and will be examined by the present work.

3 Analysis of room responses

3.1 Standard deviation of the spectral response across frequency

Let $h_p(k)$ be the discrete-time room impulse response at ''p'' different positions (e.g. obtained by varying the source and receiver positions) and let $H_p(k)$, be the corresponding N-point DFT. $H_p(k)$ calculated in decibels (dB) gives:

$$P_p(k) = 20\log(|H_p(k)|) \qquad (4)$$

In order to analyze the above data in sub-bands, the frequency range can be divided into fractional-octave bandwidths. In practice, unequal bandwidths are traditionally employed in most audio-acoustic applications, conforming to octave fractions of 1/3, 1/6,…1/64. For a specific position ''p'' and sub-band ''sb'', the standard deviation $\sigma_{p,sb}$ of the magnitude of the spectrum $P_p(k)$ for each sub-band that has a length of N frequency bins, can be calculated as:

$$\sigma_{p,sb} = \left(\frac{1}{N-1} \sum_{k=1}^{N} \left(P_{p,sb}(k) - \overline{P}_{p,sb}(k) \right)^2 \right)^{\frac{1}{2}}, \quad (5)$$

where k is the frequency index, $P_{p,sb}$ is the magnitude of the spectrum for a certain position ''p'' and a given sub-band ''sb'', and the mean value of $P_{p,sb}$ is equal to:

$$\overline{P}_{p,sb}(k) = \frac{1}{N} \sum_{k=1}^{N} P_{p,sb}(k) \qquad (6)$$

3.2 Standard deviation of the spectral response across space

The distance between the source and receiver for any impulse response $h_p(k)$ measured at position ''p'' is normalized according to the room critical distance [11],

$$d_{cr} \approx \frac{1}{4}\left(\frac{S\overline{a}}{\pi} \right)^{1/2} = \frac{1}{4}\left(\frac{cV}{\pi RT} \right)^{1/2} \qquad (7)$$

where V (m^3) is the volume of the room, c is a mathematical constant equal to 0.161, and RT (sec) is the reverberation time of the room. Expressing as P_{p,d_norm} the RTF magnitude at position p corresponding to a source − receiver distance d_{norm}, normalized according to equation (7), the standard deviation σ_{p,d_cr} can be evaluated (as in equation (5)). From this analysis, the relationship between RTF standard deviation and source-receiver distance will be derived.

3.3 Complex smoothing

Similar statistical quantities can be calculated for smoothed versions of the transfer functions. The RTF complex smoothing operation [6, 7] has been defined as:

$$H_{p,sm}(k) = \sum_{i=0}^{N-1} H_p((k-i) \bmod N) W_{sm}(m,i) \qquad (8)$$

where k is the discrete frequency index ($0 \le k \le N-1$) and $W_{sm}(m,i)$ is a low-pass filter function where m is the sample index corresponding to the cut-off frequency f_c (Hz), according to the expression $m = (N/f_s) \cdot f_c$.

Different smoothing profiles (e.g. fractional-octave smoothing, non-uniform frequency smoothing) can be used according to the needs of the applications. Typical applications of complex smoothing include real-time room compensation and dereverberation.

3.4 Histograms of reverberant signals

In order to visualize some signal-dependent statistical properties of room reverberation, as they are imposed on reverberant signals, histograms can be used. Considering an anechoic music or speech signal $s_{an}(n)$ relayed in a room that is described by an impulse response $h_p(n)$, the reverberant signal $s_{rev}(n)$ can be described as:

$$s_{rev}(n) = s_{an}(n) * h_p(n) \qquad (9)$$

The long-term magnitude spectra of the above signals ($S_{rev}(k)$, $S_{an}(k)$) normalized and expressed in dB reference to 0dB-FS (i.e. Full Scale), were analysed with the help of histograms presenting the frequency of occurrence of different spectral values.

4 Experimental procedure

The quantities described in section 3 were calculated for RTFs obtained from measurements and simulations. For the measurements the Dirac (B&K) software was employed, whereas for the simulations the ODEON (B&K) software was used. Measurements were taken in several rooms and their properties are presented in table 1.

Index	Room	RT (sec)	Volume (m^3)	Critical Distance (m)	f_{Schr} (Hz)
T 1	Theatre 1	0.75	7472	5.63	20
T 2	Theatre 2	1.00	1950	2.5	45
R 1	Room 1	0.94	1189	2.0	56
R 2	Room 2	1.1	190	0.7	152
L 1	Listening Room	0.36	95	0.9	123
S 1	Sports hall	6.4	34776	4.2	27

Table 1. Properties of the rooms where measurements were carried out.

5   Results

5.1   Fractional octave RTF analysis

The standard deviation of the magnitude of the spectrum, using 1/3 octave band analysis, is calculated for several RTFs following the procedure described in section 3.1. The impulse responses have been measured at random positions in several rooms with properties that can be found in table 1. The distance between the source and the receiver was always larger than the corresponding room critical distance and the results are presented in figure 1. A line indicating the 5.6 dB value derived by Schroeder [1] is also plotted.
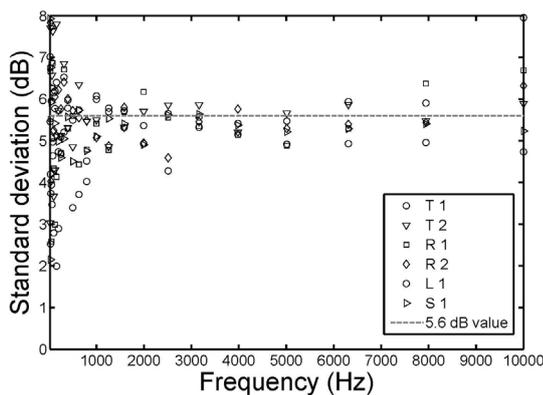
Figure 1. Standard deviation vs. frequency for RTFs (1/3 octave analysis) for random positions in several rooms.

It can be noted that the standard deviation for the frequency bands above approx. 800 Hz gives values close to Schroeder's predicted value of 5.6 dB with maximum deviation of ±1.5 dB. It is interesting to note that this trend appears irrespective of the position and the room's acoustical properties. However, for the frequency bands with frequencies lower than 800 Hz, the standard deviation seems to vary considerably with respect to the predicted value. The lower measured values are somehow expected for such low frequency bands, using 1/3 octave band analysis, since such narrow bandwidths lead to underestimation for the standard deviation, because the spectral values within a narrow band appear to be highly correlated. Additionally, below Schroeder's frequency an ambiguity in the results was also expected, for the reasons described in section 2.1.

Using the same approach, the standard deviation can be calculated and plotted, when employing 1/16 and 1/64 fractional octave analysis and the results are shown in figure 2. As expected, the higher the resolution used, i.e. the narrower the bands where the standard deviation is calculated, the lower is the measured standard deviation. This is evident from figure 2, where the values of the standard deviation are plotted as a function of frequency for a specific position in room R 2 (see table 1) using different fractional octave band analysis.

In figure 3 the standard deviation across space is calculated and plotted, following the procedure described in section 3.2, using 1/3, 1/16 and 1/64 analysis. The results correspond to the frequency range between 300 - 10000 Hz and as can be seen, at distances very close to the source the standard deviation is very low. As the distance between source and receiver increases, the standard deviation also

increases, and converges to a specific value (depending on the analysis being used) for distances higher than that of the critical distance.
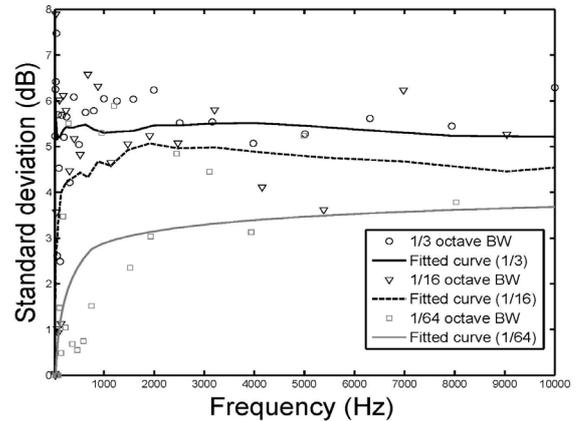
Figure 2. Standard deviation vs. frequency for RTF (1/3, 1/16 and 1/64 octave analysis) for a specific position in room R 2.

The results that appear in figure 3 are in accordance with the results obtained by Jetzt [10]. Furthermore, beyond the room critical distance, RTF standard deviation statistics across frequency and across space appear to converge to identical values depending mainly on the analysis bandwidth.
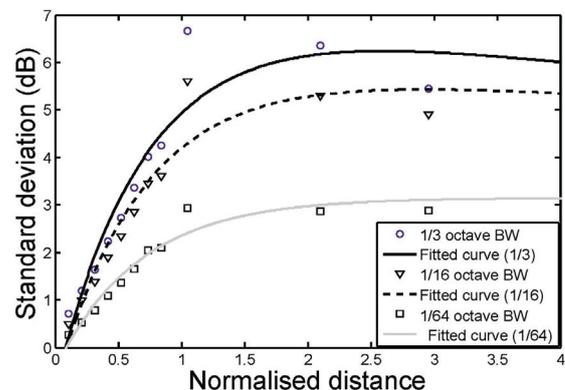
Figure 3. Standard deviation vs. normalized distance for RTFs (1/3, 1/16 and 1/64 octave analysis) taken at several positions in room R 2.

5.2   Complex smoothed fractional octave RTF responses

The same analysis as presented in section 5.1 can be obtained for complex smoothed RTFs.

In figure 4 the standard deviation, as a function of frequency for complex smoothed RTFs, obtained at random positions in several rooms (see table 1) is plotted. It can be seen that in this case the values of the standard deviation are much lower. For frequencies between 500 Hz and 7000 Hz the standard deviation appears to be lower than 2 dB, while in frequencies above 8000 Hz the values are higher for some of the rooms under examination. This was expected as complex smoothing employs averaging on the magnitude, while conforming to perceptual rules.

In figure 5 can be seen, the dependence on the analysis being used for the complex smoothed RTF standard deviation at a specific position in room R 1. Again, the higher the resolution, the lower the standard deviation.

Finally, in figure 6 the standard deviation, as a function of the normalised distance is presented. Although the behavior of the curves is similar to that appearing in figure 3, the values are now much lower, due to complex smoothing. The effect of choosing different fractional bandwidth is again evident.
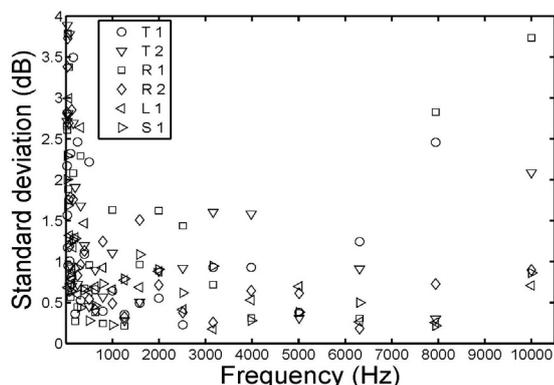
Figure 4. Standard deviation vs. frequency for complex smoothed RTFs (1/3 octave analysis) taken at random positions in several rooms.
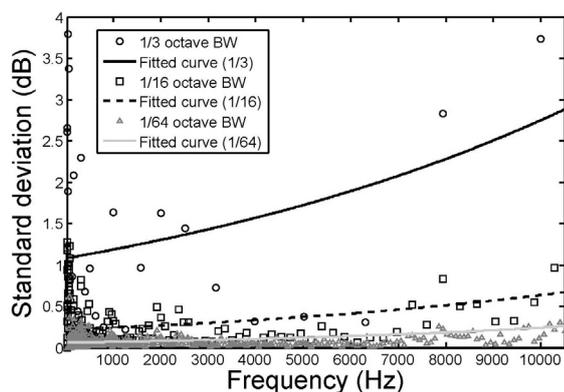
Figure 5. Standard deviation vs. frequency for complex smoothed RTF (1/3, 1/16 and 1/64 octave analysis) taken at a specific position in room R 1.
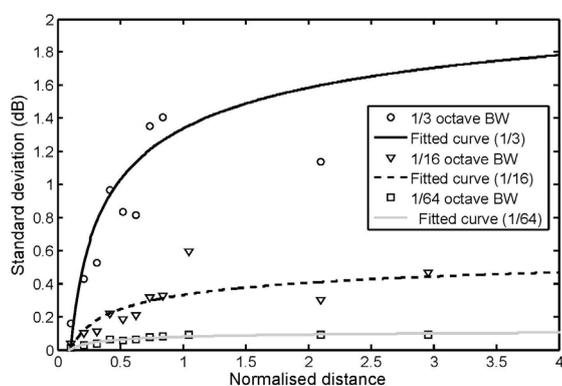
Figure 6. Standard deviation vs. normalized distance for complex smoothed RTFs (1/3 octave analysis) taken at several positions in room R 1.

5.3   Reverberant signal statistics

In figure 7, histograms are plotted presenting the statistical characteristics of the magnitude of the spectrum of two RTFs, measured at two different positions in room T 1. It can be noted that the statistics of the two RTFs follow a similar pattern, as both of them present two peaks at approximately -50 dB and -25 dB. Additionally, the histograms of the spectra of an anechoic signal and of two reverberant signals (calculated according to equation (9)) are also plotted. From these results, it appears that RTF histogram trends are superimposed on the anechoic signal histogram, leading to both a spread towards smaller values (more spectral dips) and towards the characteristic dominance of the two RTF peaks.

In figure 8 histograms showing the magnitude spectrum statistics of two other types of anechoic signals (speech and orchestral) are plotted, noting that originally the shape of their respective histograms differs significantly.
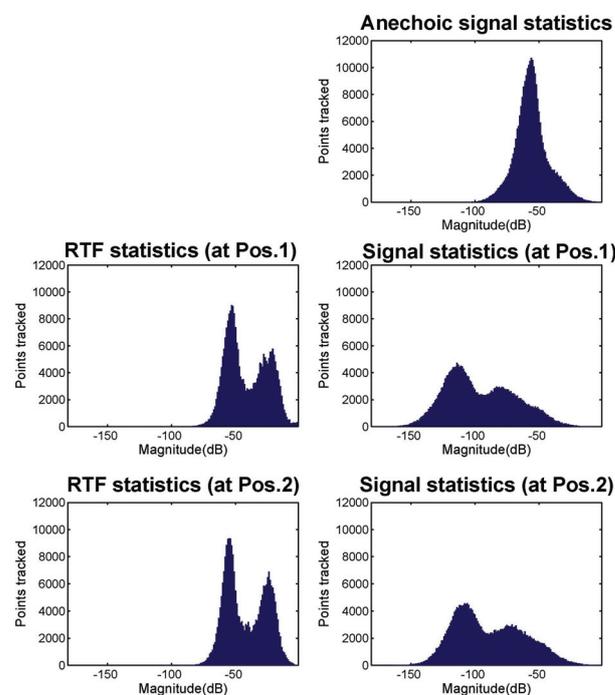
Figure 7. Histograms of the spectrum magnitude of the RTFs at two different positions in room T 1 (left column) and reverberant signal statistics (piano) when the signal is reproduced at the same positions. On the top right plot appear the statistics of the anechoic piano signal.

These spectra of the anechoic signals are then convolved with the room response obtained at a specific position in room T 1 and the statistics of the RTF of the resulting reverberant signals are also plotted on the right side of figure 8. It can be observed that the resulting reverberant signal histograms, are dominated by the characteristic 2-peak pattern due to the RTF histogram, indicating that they were reproduced in the same room.
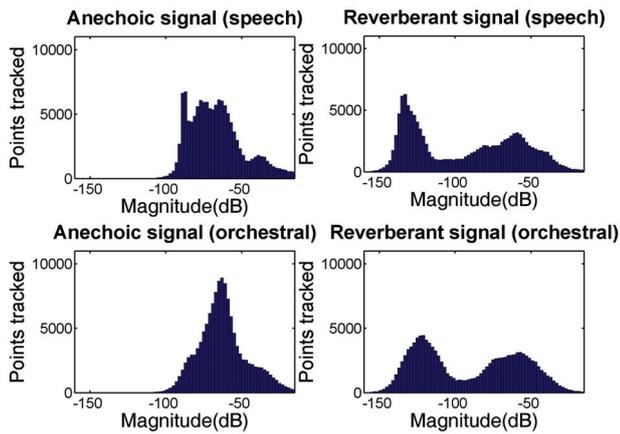
Figure 8. Histograms of the spectrum magnitude of different types of anechoic signals and of the reverberant signals, when they are reproduced at a specific position in room T1.

6 Conclusions

Statistical properties of RTFs have been investigated and the standard deviation as a function of frequency is shown to equal approximately 5.6 dB when using 1/3 fractional octave band analysis for frequencies higher than approx. 800 Hz. For lower frequencies an ambiguity of the standard deviation value exists. Furthermore, the necessity of having a sufficiently broad bandwidth in order to obtain values close to the expected value of 5.6 dB is shown. Moreover the relationship between RTF standard deviation and source-receiver distance has been also presented and it is shown that beyond room critical distance, RTF deviation approaches the result obtained for the deviation across frequency. The dependence of the values for the standard deviation on the fraction of the analysis that is used, has been also discussed.

Furthermore, the standard deviation has been also examined for complex smoothed responses, as this might be a useful approach to obtain simplified versions of RTFs that conform to perceptual rules. The obtained values for the standard deviation are shown to be much lower, biased further towards even lower values in the perceptually significant mid-frequency range and the effect of choosing different fractional bandwidth was again evident.

The RTF magnitude spectrum statistical properties and signals has been also visualized with the help of histograms. Similarities have been observed between the RTF magnitude histograms within the same room and of any reverberant signal when it is reproduced in the same room. Moreover, similarities on the magnitude statistics of different types of reverberant signals when reproduced at a specific position in a room have been noted, although the original anechoic signals presented significant differences. This approach might provide a useful framework for dereverberation techniques as it can offer information about the frequency of occurrence of the spectral values of reverberant signals which can be related to the acoustical properties and dimensions of the rooms.

The above observations may form a basis of RTF modeling based on statistical considerations. In future work, the properties of the phase and group delay of RTFs, will be investigated using a similar approach.

Acknowledgements

References

[1]   Schroeder, M. "Die statistischen Parameter der Frequenzkurven von grossen Räumen", Acustica 4: pp. 594-600 (1954). English translation in "Statistical parameters of the frequency response curves of large rooms", Journal of the Audio Engineering Society, Vol. 35, No. 5, pp. 299–306 (1987).

[2]   Polack, J.-D. "La transmission de l'énergie sonore dans les salles". Thése de Doctorat d' Etat, Universitairé du Maine, Le Mans (1988)

[3]   Polack, J.-D. "Playing billiards in the concert hall: The mathematical foundation of geometrical room acoustics". Applied Acoustics, Vol. 38, p. 235-244 (1993)

[4]   Jean-Marc Jot, Laurent Cerveau, and Olivier Warusfel, "Analysis and Synthesis of Room Reverberation Based on a Statistical Time-Frequency Model", Proc. of the 103rd AES Conv., preprint 4629 (1997)

[5]   Usher J. and Benesty J., "Enhancement of Spatial Sound Quality: A New Reverberation-Extraction Audio Mixer", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 15, No. 7 (2007)

[6]   P. Hatziantoniou and J. Mourjopoulos, "Generalised Fractional-Octave Smoothing of Audio and Acoustic Responses", J. Audio Eng. Soc., Vol. 48, No. 4, pp. 259-280 (2000)

[7]   P. Hatziantoniou and J. Mourjopoulos, "The application of complex spectral smoothing in the analysis of room acoustics.", ASME-Greek Section, Proceedings of First Nat. Conf. on Recent Advances in Mech. Eng (2001)

[8]   Schroeder, M. R. "The 'Schroeder frequency' revisited" (55). Journal of the Acoustical Society of America, Vol. 99, No. 5, pp. 3240 (1996)

[9]   Diestel H. G. "Zur Schallausbreitung in Reflexionsarmen Raumen", Acustica 12, 113-118 (1962)

[10]   Jetzt, J. "Critical distance measurement of rooms from the sound energy spectral response" Journal of the Acoustical Society of America, Vol. 65, No. 5, pp. 1204–11 (1979)

[11]   Kuttruff H., Room Acoustics, 3rd ed. Essex, U.K.: Elsevier (1991)