Audio Engineering Society

Convention Paper 8423

Presented at the 130th Convention
2011 May 13–16 London, UK

Statistical Analysis of Binaural Room Impulse Responses

Eleftheria Georganti[1], Alexandros Tsilfidis[1], and John Mourjopoulos[1]

[1] *Audio and Acoustic Technology Group, Dept. of Electrical and Computer Engineering, 26504, University of Patras, Greece*

Correspondence should be addressed to Eleftheria Georganti (egeorganti@upatras.gr)

ABSTRACT

In previous work of the authors, the spectral magnitude of room transfer functions (RTFs) was analyzed using histograms and statistical quantities (moments), such as the kurtosis and skewness. In this work, the above analysis is extended to binaural room impulse responses (BRIRs) and the dependence of the statistical measures on the room acoustical properties, such as the reverberation time, the room size and the source-receiver distance is discussed. Emphasis is given on the binaural measure of the magnitude squared coherence (MSC) which is considered to be an important cue for binaural hearing related to perceptual aspects such as the source width, the envelopment and the spaciousness. After a brief overview of the existing MSC models, a perceptually-motivated MSC implementation is examined, based on a gammatone filterbank. MSC results for various rooms and source-receiver positions are presented and related to the existing MSC models.

1. INTRODUCTION

The analysis of single channel room impulse responses (RIRs) and especially the analysis of binaural room impulse responses (BRIRs) highlights perceptually significant features of room acoustics, given that the early reflections or the late components of the responses contribute to the characteristics and the acoustic quality of a hall. The amplitude, the direction and the arrival time of each discrete reflection may provide an insight into the details of the acoustic behaviour of the room and contribute to the overall acoustical quality. However, the information contained in each early reflection component as well as in the late part of the impulse response is not interpreted individually by the human auditory system since global perceptual cues and measures are employed by the listeners. Statistical analysis of these functions assists in the interpretation of such perceptual cues and for many years such approaches have been a valuable tool in analysing RIRs [1–8] or BRIRs [9, 10]. Such statistical models can lead to RIRs/BRIRs simplifications that have been proved to be useful for several audio applications, such as room acoustic optimization, room modeling, auralization and dereverberation techniques.

In [11–13], the spectral magnitude of RTFs is analyzed using histograms and statistical moments. In this work, the above analysis is extended to binaural room impulse responses (BRIRs) by calculating these statistical measures for the binaural cues and showing their dependence on the room acoustical properties, such as the reverberation time, the size of the room and the source-receiver distance.

Typical interaural measures used for the analysis of both binaural signals and room impulse responses are (i) the interaural level difference (ILD), (ii) the interaural time difference (ITD) and (iii) the interaural coherence (IC). ILD is the level difference between the sound arriving at the two ears, ITD is the difference in the arrival time of a sound between the two ears and IC is basically the strength of correlation between the left and right ear signals.

This work mainly focuses on the IC and more specifically on the magnitude squared coherence (MSC). The MSC is examined from a perceptual point of view by processing the BRIRs via a gammatone filterbank for various rooms and source-receiver distances and the results are discussed with respect to the existing MSC models. The results indicate that when the diffuse reverberant sound field assumptions do not apply, such an analysis may provide a promising framework for the extraction of novel coherence models.

This paper is organised as follows: In Section 2, the properties of the magnitude squared coherence are discussed, an overview of existing models is given and a perceptually-compliant calculation method for the MSC is proposed. Moreover, the statistical measures that are used in this study are introduced. In Section 3, the results of the proposed MSC analysis and the corresponding statistical measures are presented and discussed. Finally, conclusions are drawn in Section 4.

2. INTERAURAL COHERENCE ESTIMATION

The interaural coherence (IC) is commonly used for studying the similarity between the left and the right binaural signals ($s_L(t)$ and $s_R(t)$ respectively):

$$IC_{S_L,S_R}(f) = \frac{|\langle S_L(f) \cdot S_R^*(f)\rangle|}{\sqrt{\langle S_L(f) * S_L^*(f)\rangle \langle S_R(f) \cdot S_R^*(f)\rangle}} \quad (1)$$

where t and f are the time and frequency indices, $S_L(f)$ is the Fourier transform of $s_L(t)$ and $S_R(f)$ is the Fourier transform of $s_R(t)$, $*$ denotes the complex conjugate and $\langle x\rangle$ is the expected value of x. The magnitude squared coherence (MSC) is referred to the square of Eq. (1).

Past research efforts have studied and modelled the diffuse field coherence. In [14], the theoretical coherence for a perfectly diffuse sound has been shown to have a sinc function behaviour:

$$R(kd) = \frac{sin(kd)}{kd} = sinc(kd) \quad (2)$$

where $k = 2\pi f/c$. The correlation length d is the microphone spacing, which marks the transition between essentially correlated and uncorrelated signals and may be defined as the first zero of $R(kd)$. Thus, $kd = \pi$ or $d = \lambda/2$. Later on, in [15] the authors have shown that the noise field coherence values are influenced by the shape of the human's head. In [16], a more detailed model is presented, showing that a rigid sphere model of the head makes the first minimum in the coherence curve appear at a lower frequency than for the spaced omni microphone model. Recently, a semi-analytical signal processing model for the binaural coherence which takes shadowing effects of the head into account has been presented [17]. This model is in agreement with the measured coherence and it has been efficiently employed in a recently developed binaural dereverberation method [18].

From this, it appears that IC modelling has been well established in the case of perfectly diffuse reverberant sound fields. However, given the wide application range potential, it might be also useful to examine the behaviour of IC in smaller rooms or in positions within rooms where the perfectly diffuse assumption does not apply. Moreover, since IC is considered to be an important cue for binaural hearing related to perceptual aspects such as the source width, the envelopment and the spaciousness [19], but also to the detection of the direct sound in reverberant environments [20], it would be also interesting to examine the IC following a more perceptual point of view.

In this study, the signals of interest, being the two channels of a BRIR, denoted as $h_L(n)$ and $h_R(n)$ are passed through a Gammatone cochlear filterbank [21] and decomposed in k critical bands. Then, each band is divided into i subsequent time frames and the IC is estimated as:

$$\Phi(i,k) = \frac{R_{LR}(i,k)}{\sqrt{R_L(i,k) \cdot R_R(i,k)}} \tag{3}$$

where

$$R_L(i,k) = a_1 \cdot \sum_{n=1}^{N} h_L^2(n) + (1-a_1) \cdot R_L(i-1,k), \tag{4}$$

$$R_R(i,k) = a_1 \cdot \sum_{n=1}^{N} h_R^2(n) + (1-a_1) \cdot R_R(i-1,k), \tag{5}$$

$$R_{LR}(i,k) = a_1 \cdot \sum_{n=1}^{N} (h_L(n) \cdot h_R(n)) + (1-a_1) \cdot R_{LR}(i-1,k), \tag{6}$$

N is the number of samples in each time frame and a_1 is a smoothing factor determined from the time constant T=10 ms and the sampling frequency f_s in Hz, given by [20],

$$a_1 = \frac{1}{T \cdot f_s}. \tag{7}$$

In this way an IC value is calculated for every successive time frame i and frequency band k. The magnitude squared coherence $\Gamma(i,k)$ is then calculated as the squared value of $\Phi(i,k)$ and the mean $\Gamma(i,k)$ for each frequency band k can be calculated as:

$$\mu_{LR}(k) = \frac{1}{\tau} \sum_{i=1}^{\tau} \Gamma(i,k) \tag{8}$$

where τ is the total number of the examined time frames.

The mean can be also calculated over the full frequency range as,

$$M_{LR} = \frac{1}{K} \sum_{k=1}^{K} \mu_{LR}(k) \tag{9}$$

where K is the total number of the frequency bands; in this case being the number of critical bands.

Using the above equations, higher order statistical measures can be calculated, i.e. the MSC skewness for each frequency band k:

$$\sigma_{LR}(k) = \frac{\frac{1}{\tau}\sum_{i=1}^{\tau}(\Gamma(i,k) - \mu_{LR}(k))^3}{\left[\frac{1}{\tau}\sum_{i=1}^{\tau}(\Gamma(i,k) - \mu_{LR}(k))^2\right]^{\frac{3}{2}}}, \tag{10}$$

and for the full frequency range as,

$$S_{LR} = \frac{\frac{1}{N}\sum_{k=1}^{K}(s_{LR}(k) - M_{LR})^3}{\left[\frac{1}{K}\sum_{i=1}^{K}(s_{LR}(k) - M_{LR})^2\right]^{\frac{3}{2}}}. \tag{11}$$

Alternatively, the kurtosis $\lambda_{LR}(k)$ in each subband can be derived as,

$$\lambda_{LR}(k) = \frac{\frac{1}{\tau}\sum_{i=1}^{\tau}(\Gamma(i,k) - \mu_{LR}(k))^4}{\left[\frac{1}{\tau}\sum_{i=1}^{\tau}(\Gamma(i,k) - \mu_{LR}(k))^2\right]^{2}}, \tag{12}$$

and for the full frequency range as,

$$\Lambda_{LR} = \frac{\frac{1}{N}\sum_{k=1}^{K}(s_{LR}(k) - M_{LR})^4}{\left[\frac{1}{K}\sum_{i=1}^{K}(s_{LR}(k) - M_{LR})^2\right]^{2}}. \tag{13}$$

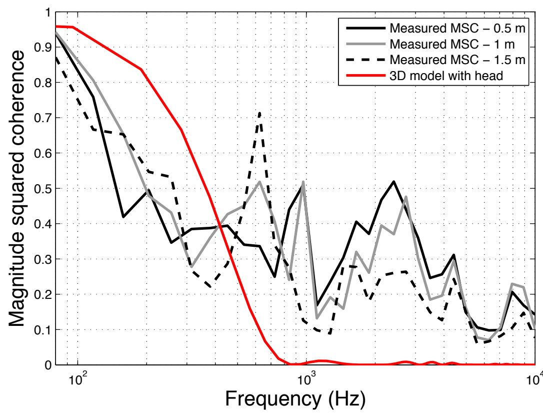
3. RESULTS

3.1. Interaural Coherence vs. frequency

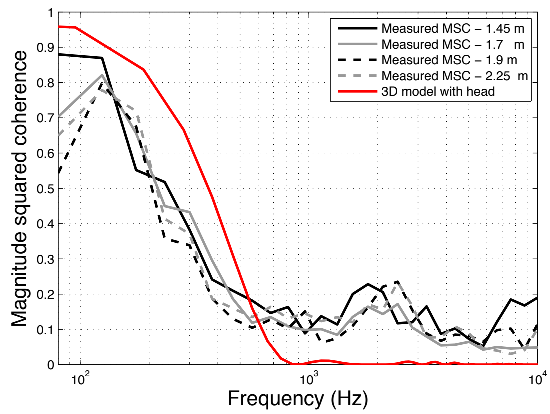
In this section, the results obtained from the analysis of BRIRs of various rooms are presented. The BRIRs were taken from the Aachen Impulse Response (AIR) database [22], which is a set of binaural impulse responses that were measured in a wide variety of rooms using a dummy head; their geometrical and acoustical characteristics can be found in Table 1. The MSC ($\Gamma(i,k)$) is calculated as described in Section 2 and the mean of each frequency band is calculated according to Eq. 8, using time frames of

Room	Volume (m^3)	RT (sec)	d_{crit} (m)
Studio booth	11.9	0.12	0.6
Meeting Room	116	0.23	1.3
Office Room	92.8	0.43	0.8
Stairway Hall	-	1.47	-

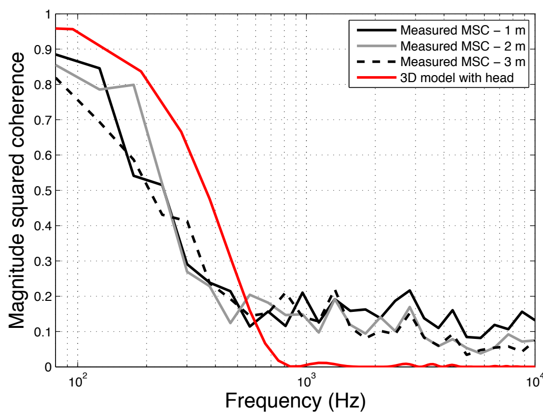
Table 1: Geometrical and acoustical characteristics of the rooms [22].



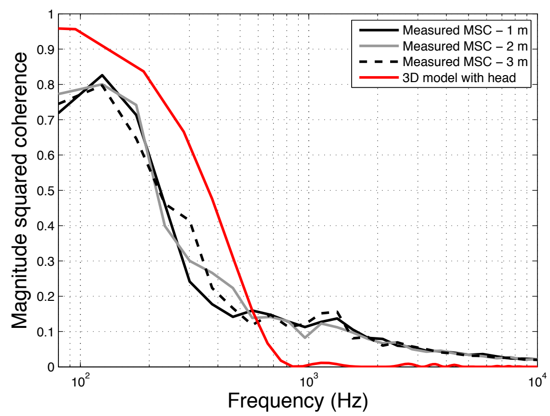
(a) Studio booth



(b) Meeting room



(c) Office room



(d) Stairway hall

Fig. 1: Magnitude squared coherence for the rooms. Their geometrical and acoustical characteristics can be found in Table 1.

40 ms duration (1920 samples at a sampling rate of 48 kHz) and an overlap of 20 ms.

3.1.1. Distance dependence

Here, the dependence of the MSC on the distance between the source and the receivers (dummy head) is examined. The results for each room and various distances can be seen in Fig. 1(a)-1(d). It can be noted that in the case of the studio booth, which is a small-sized room (see Fig. 1(a)) the MSC curve presents prominent peaks in the frequency region between 1 kHz and 5 kHz and a lower peak around 8 kHz. These high MSC values imply that the left and right BRIRs are correlated in these frequency regions. The above MSC trend is significantly different from the diffuse-field expected curve [17] (red line) since this studio booth is a small room with very low reverberation time (0.12 s) and in such room the diffuse reverberant sound field theory does not apply. Moreover, it can be noted that the peaks on the MSC appear in the same frequency values independently of the distance between source and receiver, implying that these peaks might be related to room acoustical properties such as early reflections or discrete modes. These peaks take lower values as the distance increases, implying that they become less correlated probably because of the relative increase in the reverberant field.

On the other hand, the MSC curve for the meeting room follows a different trend (see Fig. 1(b)), although it has also a low reverberation time of similar range with that of the studio booth. In the meeting room, all the measurements are taken for distances beyond 1.4 m; peaks appear at the MSC curves, but presenting lower values than before, and not in the same frequency bins across all the positions. Again, the BRIR measured at the closest position between source and receiver presents the highest coherence values across frequency.

In Fig. 1(c), the MSC curve for the office room is plotted as a function of frequency. It appears to be more smoothed and several peaks appear again in the frequency range between 1 kHz and 5 kHz, presenting the highest MSC values at the closest distance. Finally, the MSC curve for the stairway hall (see Fig. 1(d)) appears more smoothed than those of the other tested rooms presenting a reasonable approximation to the diffuse field theory assumptions.

3.1.2. Angle dependence

Here, the MSC is examined as a function of the azimuth angle. In Fig. 2, the MSC for the lecture room is plotted for various angle orientations, from 0° to 180° with a step of 15° ([0° : 15° : 180°]) having a distance of 2 m between source and receiver. It can be noted that the curves present similar trends, independently of the angle orientation, given the theoretically uniform distribution of the reflection arrival angles in such a largely diffuse reverberant room.

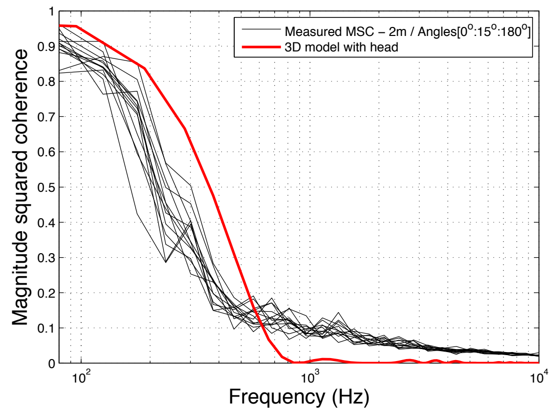
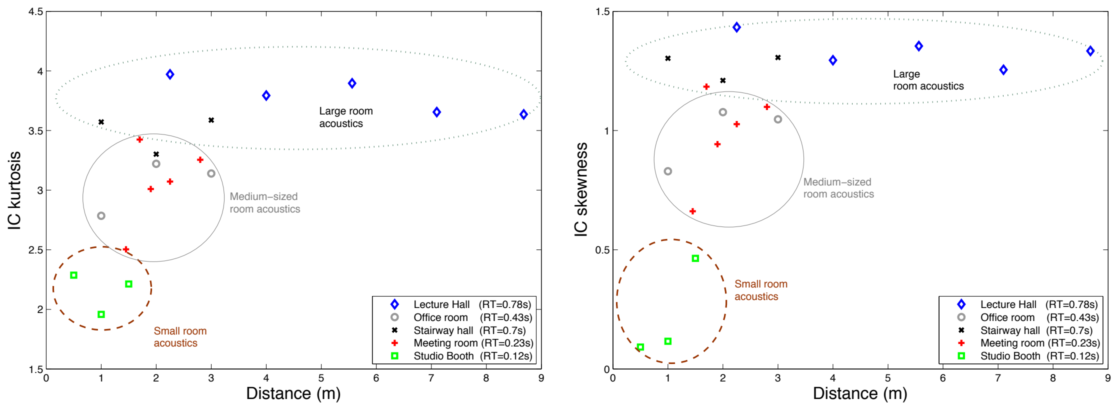


Fig. 2: Magnitude squared interaural coherence for the lecture room for various angle orientations [0° : 15° : 180°] having distance between source and receiver of 2 m.

3.2. Statistical moments vs. distance

In this section, the statistical moments of the kurtosis and skewness are examined over the full frequency range and over all the time frames as a function of distance. The statistical moments are calculated according to Eq. (11) and (13). Skewness is a measure of the asymmetry of the probability distribution of a random variable. When the skewness value is positive, the right "tail" of the histogram is longer; most of the values of the distribution would be concentrated on the left of the figure and relatively few high values appear, making this effect more evident for higher skewness values. From Fig. 3(a) it can be seen that as the distance increases the skewness value also increases, implying that the MSC takes lower values and as it has been also shown in Fig. 1, where small distances lead to higher MSC values. Kurtosis is a measure of the "peakedness" of the probability distribution



(a) MSC skewness (b) MSC kurtosis

Fig. 3: Statistical moments as a function of distance.

of a random variable and a high kurtosis value indicates that the distribution has a sharper peak, while a low kurtosis distribution has a more spread peak. The results for the MSC kurtosis as a function of distance can be seen in Fig. 3(b).

It can be noted that both curves (see Fig. 3(a) and 3(b)) follow similar trends, increasing as a function of distance, presenting some convergence for further distances. Skewness converges to a value around 1.3 and kurtosis to 3.7. This convergence is related to the assumptions of the diffuse sound field theory, where the statistical measures of the sound field would present similar behaviour (such as the 5.6 dB value of the standard deviation [1]). Furthermore, it can be also observed that results derived in rooms dominated by "small room acoustics" discrete reflection sound field are grouped in regions of low skewness and kurtosis values (studio booth). In-between the "large and small room acoustics" regions, a medium-sized room acoustics region can be defined according to the skewness and kurtosis values, which in this case includes the positions within the office and meeting rooms.

4. CONCLUSIONS

In this work, the perceptually-compliant binaural measure of the magnitude squared coherence (MSC) has been studied and the results have been compared to existing IC models. Higher statistical moments of the MSC of BRIRs have been calculated and their dependence on the room properties/positions has been underlined. These preliminary results show the potential for developing a generalized framework incorporating novel coherence models for positions within rooms where the diffuse sound field theory assumptions for reverberation do not apply, while conforming to perceptual rules.

Further work should include the mathematical framework of the above relationships as it can be seen that there is correlation of the MSC curves to the direct to reverberant energy ratio of the BRIRs. Similar analysis should be applied directly to BRIRs signals and examine if the above relationships and findings are still valid. Such analysis can lead to more robust room dependent models that could assist various audio applications such as dereverberation, speech enhancement.

ACKNOWLEDGEMENTS

The authors would like to thank Tobias May from the University of Oldenburg for providing the Matlab code for the binaural framework. The research activities that led to these results, were co-financed by Hellenic Funds and by the European Regional Development Fund (ERDF) under the Hellenic National Strategic Reference Framework (ESPA) 2007-2013, according to Contract No. MICRO2-38/E-II-A.

5. REFERENCES

[1] M. Schroeder. Statistical parameters of the fre-

quency response curves of large rooms. *J. Audio Eng. Soc.*, 35(5):299–306, 1987.

[2] M. R. Schroeder. Natural sounding artificial reverberation. *J. Audio Eng. Soc*, 10(3):219–223, 1962.

[3] M. R. Schroeder and K. H. Kuttruff. On frequency response curves in rooms; comparison of experimental, theoretical, and monte carlo results for the average frequency spacing between maxima. *The Journal of the Acoustical Society of America*, 34(1):76–80, 1962.

[4] M. R. Schroeder. Frequency-correlation functions of frequency responses in rooms. *The Journal of the Acoustical Society of America*, 34(12):1819–1823, 1962.

[5] J-D. Polack. *La transmission de l'énergie sonore dans les salles*. PhD thesis, Le Mans, France, 1988.

[6] J. J. Jetzt. Critical distance measurement of rooms from the sound energy spectral response. *J. Acoust. Soc. Amer.*, 61(S1):S34–S34, 1977.

[7] R. V. Waterhouse. Statistical properties of reverberant sound fields. *J. Acoust. Soc. Amer.*, 43(6):1436–1444, 1968.

[8] D. Lubman. Fluctuations of sound with position in a reverberant room. *J. Acoust. Soc. Amer.*, 44(6):1491–1502, 1968.

[9] Fritz Menzer. *Binaural Audio Signal Processing Using Interaural Coherence Matching*. PhD thesis, Lausanne, 2010.

[10] F. Menzer and C. Faller. Investigations on an early-reflection-free model for brirs. *J. Audio Eng. Soc*, 58(9):709–723, 2010.

[11] E. Georganti, J. Mourjopoulos, and F. Jacobsen. Analysis of room transfer function and reverberant signal statistics. In *Acoustics '08*, Paris, France, 2008.

[12] E. Georganti, T. Zarouchas, and J. Mourjopoulos. Reverberation analysis via response and signal statistics. In *Audio Engineering Society Convention 128*, 2010.

[13] E. Georganti and J. Mourjopoulos. Statistical relationships of room transfer functions and signals. *submitted to Forum Acusticum*, 2011.

[14] R. K. Cook, R. V. Waterhouse, R. D. Berendt, S. Edelman, and M. C. Thompson. Measurement of correlation coefficients in reverberant sound fields. *The Journal of the Acoustical Society of America*, 27(6):1072–1077, 1955.

[15] I. M. Lindevald and A. H. Benade. Two-ear correlation in the statistical sound fields of rooms. *The Journal of the Acoustical Society of America*, 80(2):661–664, 1986.

[16] A. Avni and B. Rafaely. Interaural cross correlation and spatial correlation in a sound field represented by spherical harmonics. In *Ambisonics Symposium*, Graz, Austria, 2009.

[17] M. Jeub, M. Dorbecker, and P. Vary. A semi-analytical model for the binaural coherence of noise fields. *Signal Processing Letters, IEEE*, 18(3):197–200, March 2011.

[18] M. Jeub, M. Schafer, T. Esch, and P. Vary. Model-based dereverberation preserving binaural cues. *Audio, Speech, and Language Processing, IEEE Transactions on*, 18(7):1732–1745, Sept. 2010.

[19] J. Blauert. *Spatial Hearing: The Psychophysics of Human Sound Localization*. The MIT Press, 1997.

[20] C. Faller and J. Merimaa. Source localization in complex listening situations: Selection of binaural cues based on interaural coherence. *The Journal of the Acoustical Society of America*, 116(5):3075–3089, 2004.

[21] R. D. Patterson, M. Allerhand, and C. Giguere. Time-domain modelling of peripheral auditory processing: A modular architecture and a software platform. *Journal of the Acoustical Society of America*, 98:1890–1894, 1994.

[22] M. Jeub, M. Schäfer, and P. Vary. A binaural room impulse response database for the evaluation of dereverberation algorithms. In *16th International Conference on Digital Signal Processing*, pages 550–554, Santorini, Greece, 2009.