

Spatial Impulse Response Assessment in Room Acoustics Auralization

P. MAŁECKI*

AGH University of Science and Technology, Department of Mechanics and Vibroacoustics,
Al. A. Mickiewicza 30, 30-059 Krakow, Poland

Ambisonic impulse responses (also known as the Spatial Impulse Responses) are getting more and more popular in many applications of room acoustics. Measuring, modeling, synthesizing, or simulating acoustics of interiors with emphasize on their spatial features are the examples. The present article shows a methodology of room acoustics auralization assessment using the ambisonics technology. A case study is shown for room acoustics capturing (measuring the Spatial Impulse Response) and reproducing with the use of a multi-channel sound system. An discussion and considerations about fidelity of such reconstruction is provided. Furthermore, a quantitative measure for the auralization quality is proposed. Also on-site experiments are described and their results presented to illustrate the theoretical considerations and outline options for future work. Obtained results show that most of the proposed quality factors meet the assumptions. After possible future modifications and additional research, the proposed methods could be used in evaluation of room acoustic auralization.

DOI: [10.12693/APhysPolA.128.A-17](https://doi.org/10.12693/APhysPolA.128.A-17)

PACS: 43.55.+p, 43.58.+z, 43.66.+y

1. Introduction

Room acoustic auralization is a very common practice in audio production [1], room acoustic design [2], and other applications [3–5]. However, there is no measurement standard or quality assessment method that could compare an actual room with its auralization or other acoustic environment. The problem of room acoustics fidelity becomes important when psychoacoustic research is carried out on sound stimuli that are inseparable from room features [6]. However, quality estimation in this case is based mostly on visual comparison of the measured and auralized IRs (Impulse Responses) without any scale or quantitative assessment.

The auralization quality estimation is also very important in room acoustics auralizations through simulations [7]. In this case, the problem is much more complex but the method proposed in this paper could be also extended onto such simulations. Design of the method for estimation of room acoustics auralization similarity to original (simulated) room would allow to improve significantly optimization of numerical methods in room acoustics simulations and auralizations.

The quality of sound signals transmission is a widely studied issue and there are available standard methods that allow to compare original signal with the transmitted or reproduced one [8–10]. These methods cannot be directly adapted for the current problem because of many reasons. First of all, directional features of the sound field are omitted and there is no direct source signal to compare with.

Research on quality of auralization is conducted also for binaural techniques [11] but the approach to the problem is slightly different in this case, so the present study focuses on multi-channel sound playback techniques.

Practically up until as late as the 21st century, most of the spatial sound playback techniques employed 2-channel stereo, the 5.1 system popular in home theaters, and there were many studies devoted to theoretical approach to high-fidelity multi-channel playback techniques [12]. Some of the methods were developed in this area but their application required too much computing power and a lot of investment.

In recent decades, there was a rapid development of techniques that assumed the exact physical reconstruction of sound field through spatial synthesis. These methods are physically limited but nevertheless, they are continuously developed and investigated [13–16].

Techniques such as the Wave Field Synthesis or ambisonics produce high quality sound field but the fidelity of room acoustics reproduction is in general excluded from the scope of related considerations. There is some research on fidelity of these systems for pure and complex signals but auralization of room acoustics is a special case of this issue. The room acoustics defines and utilizes a large number of numerical parameters [17, 18], many standards [19], and a lot of research is conducted on room acoustics evaluation [20–24].

The basic idea of the present study is to find a way to compare the actual reverberant room acoustics with its auralization in laboratory conditions. In Sect. 2 of this paper, a general concept concerning auralization quality estimation is presented. Section 3 proposes the concept of first reflection reproduction, while Sect. 4 shows the estimation of reverberant tail reproduction. Further, Sect. 5 shows results of the experiment by means of which the

*e-mail: pawel.malecki@agh.edu.pl

proposed method was validated. Section 6 presents conclusions as well as perspectives for future research.

2. The concept of quality assessment

Currently, the IR of a room is used as a standard in room acoustic measurements methodology [19], but it is also getting more and more popular as an auralization tool for architectural purposes [25]. A typical room IR (measured with a standard omni-directional microphone) neglects spatial features of room acoustics in some way. There are only a few parameters in room acoustics that consider spaciousness which require additional measurement tools (such as a dummy head or a figure-of-8 microphone). More versatile measurement possibilities are offered by the SIR (Spatial Impulse Response), especially in the context of spatiality. SIR is a 4-channel IR matrix measured with a first-order ambisonics microphone [3].

The general assumption adopted in this study is to perform SIR measurements in the original room (which is to be auralized) and then measure the SIR of auralization. The SIR of auralization can be obtained by applying the inverse Fourier transform to quotient (Eq. (1)) of test signals spectra. The first of these test signals $s(n)$ is a standard sweep sine of fixed duration and the second one $s_{RW}(n)$ is the same test signal convoluted with original SIR and played back through the auralization system that is to be evaluated. Signals $w(n)$, $x(n)$, $y(n)$, $z(n)$ correspond to first-order ambisonic components. The “original” SIR components and the related factors are denoted with subscript “0” (zero), and the “auralized” quantities are marked with subscript “A”. All of the considered signals are calculated in the digital domain so n in further formulae refers to the signal sample, f denotes the discrete frequency, and F^{-1} is the inverse discrete Fourier transform:

$$w_A(n) = F^{-1} \left[\frac{s_{RW}(f)}{s(f)} \right], \quad (1)$$

In Eq. (1), $s_{RW}(n)$ is the same signal as $s_W(n)$, but simultaneously played and recorded by the room auralization system:

$$s_{RW}(n) = w_A(n) * s(n), \quad (2)$$

and $s_W(n)$ is straightforward calculated as:

$$s_W(n) = w_0(n) * s(n). \quad (3)$$

The resultant SIRs, referred to in the following as the original SIR and the auralized SIR, respectively, should be compared in order to find to what extent they differ from each other. The comparison should take into account properties of the human auditory system.

2.1. Determining of IR division time intervals

First of all, the proposed method involves dividing the obtained SIRs into two parts and comparing each of the parts separately. In general, a room IR can be divided into an early part, where the perception is influenced mostly by discrete reflections, and a late part, which is rather stochastic. Most of commonly used room acoustics parameters such as Clarity, Definition, Support, etc. [17]

also divide IR into the early and the late part, but defined within arbitrarily selected time intervals. Such selection is obviously based on the specific purpose for which the parameter subject to division is used, but for the current application it is proposed to analyze the Spatial Impulse Response on the basis of the room decay curve. The early part of SIR corresponds to the time interval $[0, \text{EDT}/6]$, where EDT is the Early Decay Time in seconds. The late, stochastic SIR part is defined for the interval $[\text{EDT}/6, \text{RT}]$, where RT is the Reverberation Time expressed in seconds as well. The quality-related factors defined in the following and concerning the early SIR part will be denoted in further text with subscript “EDT”, while those relating to the late part will be marked with subscript “st”.

3. Discrete reflection analysis in the early SIR part

To obtain discrete reflections from the early part of SIR measured by an ambisonic microphone, the sound field diffuseness is calculated according to [3]. Results of calculations show that detection of particular significant discrete reflections is very ambiguous. Energy ratios of direct and diffuse reflections are too low and vary with room properties.

After a series of preliminary measurements and an analysis, it has been decided to apply time windowing with $\Delta t = 0.5$ ms time constant. This corresponds to the wavelength of approximately 17 cm and ensures quite good spatial resolution but limits the frequency range. Analysis of the early part is based on the energy amplitude, while discussion concerning the late part is focused on spectral information. For each of the timeframes, spherical coordinates φ and θ of estimated sound intensity are calculated using Eqs. (4) and (5), where atan2 refers to numerical implementation of the inverse tangent function. The function atan2 takes into account signs of function arguments to determine the corresponding angle quadrant.

$$\varphi_i = \text{atan2} \left(\frac{\sum_{n=i}^{i+\Delta t} \left(x(n)^2 \text{sign}x(n) \text{sign}w(n) \right)}{\sum_{n=i}^{i+\Delta t} \left(y(n)^2 \text{sign}y(n) \text{sign}w(n) \right)} \right) \quad (4)$$

for $0 < i \leq \frac{\text{EDT}}{\Delta t}$,

$$\theta_i = \text{atan2} \left(\frac{\sum_{n=i}^{i+\Delta t} \left(z(n)^2 \text{sign}z(n) \text{sign}w(n) \right)}{\sum_{n=i}^{i+\Delta t} \left(\sqrt{x(n)^2 + y(n)^2} \right)} \right) \quad (5)$$

for $0 < i \leq \frac{\text{EDT}}{\Delta t}$.

All the values are calculated for both “original” and “auralized” SIR. The “original” SIR is the actual room B-format impulse response, whereas the “auralized” SIR is this obtained from the sound space reconstructed by an loudspeaker array.

For the early SIR part, the factors proposed to be used to estimate the auralization quality are the maximum values of cross-correlation function for spherical coordinates series:

$$R_{\varphi_{\text{EDT}}} = \max |R_{\varphi_0 \varphi_A}|, \quad (6)$$

$$R_{\theta_{\text{EDT}}} = \max |R_{\theta_0 \theta_A}|. \quad (7)$$

In order to take into account the psychoacoustic localization characteristic for human perception [26], the following additional simple factors are proposed:

$$F_{\varphi_{\text{EDT}}} = \frac{1}{N} \sum_{i=1}^N \begin{cases} 1 \wedge_{-45^\circ \leq \varphi_0 \leq 45^\circ} \leftrightarrow |\varphi_0 - \varphi_A| \leq 5^\circ, \\ 1 \wedge_{-45^\circ \geq \varphi_0 \geq 45^\circ} \leftrightarrow |\varphi_0 - \varphi_A| \leq 10^\circ, \\ 0 \text{ otherwise} \end{cases} \quad (8)$$

$$F_{\theta_{\text{EDT}}} = \frac{1}{N} \sum_{i=1}^N \begin{cases} 1 \leftrightarrow |\theta_0 - \theta_A| \leq 10^\circ, \\ 0 \text{ otherwise.} \end{cases} \quad (9)$$

The symbol N denotes the total number of spherical coordinates corresponding to early parts of SIRs. To determine similarity in the amplitude domain, acoustic pressure values for both SIRs are compared. The maximum value of normalized cross-correlation function of pressure components for early SIR parts is calculated for this purpose:

$$R_{\text{EDT}} = \max |R_{w_0 w_A}|. \quad (10)$$

4. Analysis of stochastic parts of SIRs

In the time interval [EDT/6, RT], SIR is analyzed in the frequency domain only. The sound intensity vector in the frequency domain is estimated by means of I_x , I_y , I_z determined according to [3]. Given intensity vector coordinates, it is possible to calculate spherical coordinates Φ and Θ for the stochastic part of SIR in the spectrum domain by using the formulae:

$$\Phi(f) = \text{atan2} \left(\frac{I_x}{I_y} \right), \quad (11)$$

$$\Theta(f) = \text{atan2} \left(\frac{I_z}{\sqrt{I_y^2 + I_x^2}} \right). \quad (12)$$

Similar factors for the late stochastic part of SIR are proposed by way of analogy of the early part. These factors will be selected in a way reflecting spectral coloration of auralized SIR. The proposed factors take into account isophonic curves as well as the sound localization properties of human hearing. According to the American National Standards Institute [27], a polynomial P_A of approximated A-weighted curve can be estimated for linear and logarithmic values with the use of the following formulae:

$$P_A(f) = \frac{12200^2 f^4}{(f^2 + 20.6^2) \sqrt{(f^2 + 107.7^2) (f^2 + 737.9^2) (f^2 + 12200^2)}}, \quad (13)$$

$$A(f) = 2.0 + 20 \log_{10} P_A(f). \quad (14)$$

Azimuth- and altitude-related similarity factors for SIRs are proposed in Eqs. (15)–(18). They are defined in analogy to early part factors but with some above-mentioned modifications:

$$R_{\Phi_{\text{st}}} = \max |R_{\Phi_0 \Phi_A}|, \quad (15)$$

$$R_{\Theta_{\text{st}}} = \max |R_{\Theta_0 \Theta_A}|, \quad (16)$$

In the following equations, symbol N refers to the total number of frequency bands of the stochastic parts of SIRs:

$$F_{\Phi_{\text{st}}} = \frac{1}{N} \sum_{i=1}^N \begin{cases} 1 \wedge_{(-45^\circ \leq \varphi_0 \leq 45^\circ)} \leftrightarrow P_A |\Phi_0 - \Phi_A| \leq 5^\circ, \\ 1 \wedge_{(-45^\circ > \varphi_0 > 45^\circ)} \leftrightarrow P_A |\Phi_0 - \Phi_A| \leq 10^\circ, \\ 0 \text{ otherwise,} \end{cases} \quad (17)$$

$$F_{\Theta_{\text{st}}} = \frac{1}{N} \sum_{i=1}^N \begin{cases} 1 \leftrightarrow P_A |\Theta_0 - \Theta_A| \leq 10^\circ, \\ 0 \text{ otherwise.} \end{cases} \quad (18)$$

Similarity of spectra of the stochastic parts of SIRs can be represented by a quantity calculated as a simple subtraction with A-weighted curve correction (Eq. (14)) of the obtained difference:

$$F_{\text{st}} = \frac{1}{N} \sum_{i=1}^N \{ ||20 \log_{10} W_0|| - ||20 \log_{10} W_A|| + A \}. \quad (19)$$

All of the proposed factors are meant to estimate similarity between auralized and original SIR measured in actual diffuse field.

5. An experiment on SIR variability

In order to find interpretation of magnitude of the proposed factors, a dedicated experiment has been designed. The idea behind the test was to check how sensitive are the proposed factors to geometry of the measurement setup. The SIR impulse response measurement was carried out in 3 different rooms. In each of the rooms, several different measurement points were selected. Each point was chosen at a fixed distance from the reference point. The following distances were used: 0 (measurement in the same place), 1 cm, 10 cm, 0.5 m, and 2 m. Calculated values of the proposed similarity factors for SIRs in these locations are shown in Figs. 1a–c. The measurements were taken in following rooms:

- a laboratory room — with floor area of 74 m² and RT = 1 s for 1 kHz (Fig. 1a);
- a lecture room with 110 seats and RT = 1.6 s for 1 kHz (Fig. 1b);
- a large church, capable to accommodate more than 3 000 people, with RT = 7.3 s for 1 kHz (Fig. 1c).

Results obtained in such geometrical experiment show how the proposed factors differ in respect of room features and distance. The exact results are shown in Fig. 1, where all of the parameters except for F_{st} are dimensionless. According to Eq. (16), F_{st} is expressed in decibels.

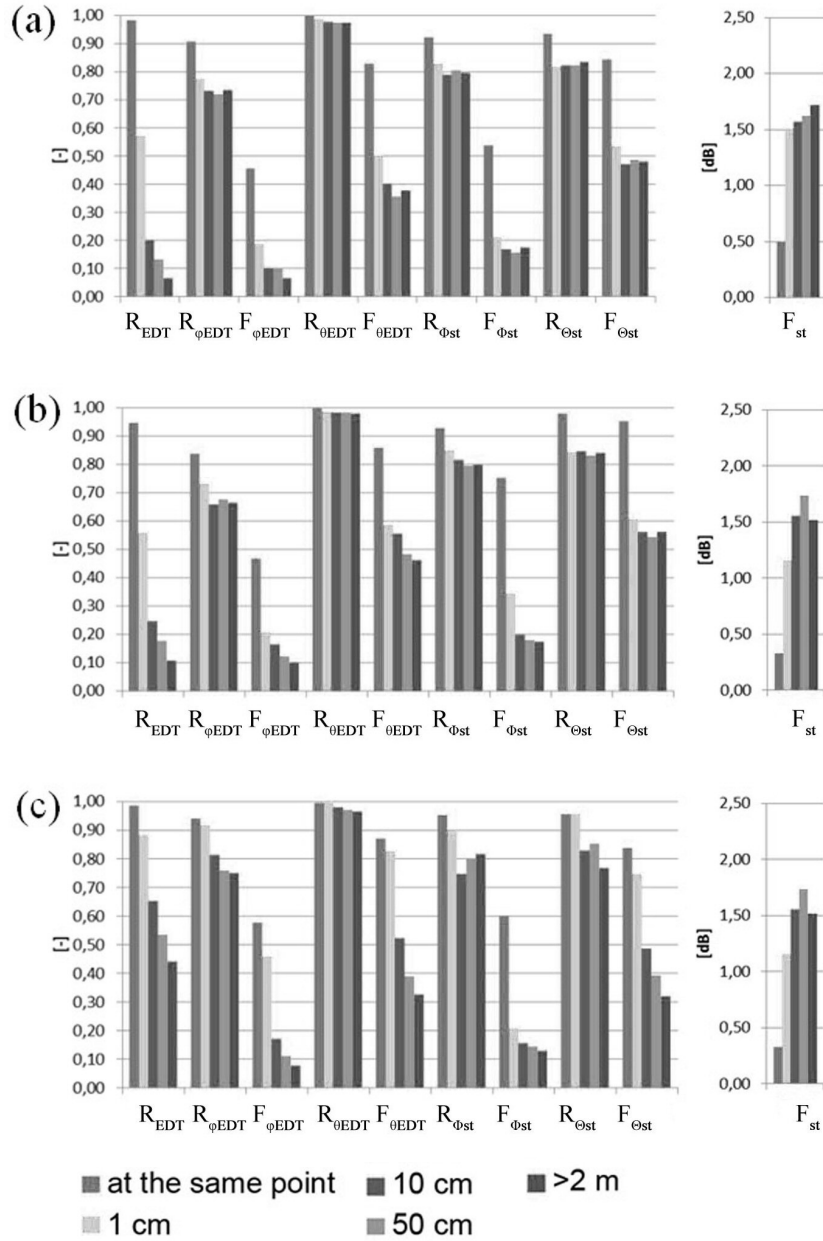


Fig. 1. Values of proposed similarity factors for 3 different rooms and 4 different measurement geometrical setups.

The most useful factors would be those values of which change proportionally to distance. Values obtained for 50 cm are considered critical because fidelity of sound field reconstruction in laboratory conditions like in same room within a 50 cm radius should be sufficient. Therefore, the differences between two SIRs measured in the same room 50 cm apart should be larger than those between SIR measured in a reference room and its synthesis obtained by means of a sound system. Some of the proposed factors, such as $R_{\phi EDT}$, $R_{\phi st}$, and $R_{\theta st}$, do not decrease with distance. Moreover, R_{EDT} varies very much depending on the room type. This could be considered an argument against usefulness of these particular factors for the purpose of room acoustics auralization quality assessment.

With changing geometrical conditions, other parameters change their values monotonically. This suggests that they are useful for the purpose of solving the problem put forward in this paper. The observed differences in values of these parameters are, in most cases, relatively large for displacements of the order of 1 cm and much smaller for other displacement values. In order to verify usefulness of the proposed parameters in a more substantiated manner, it seems to be appropriate to check how these parameters would vary as a result of an additional distortion of the impulse response introduced in a controlled way with respect to its original form or to compare it with the response obtained by means of synthesis. The expected result of such a comparison would consist in differences between modified or synthesised response

functions being much larger than those observed between different responses registered within the same room.

6. Conclusions and summary

The concept of introducing a set of auralization quality factors (AQFs) is presented in this paper. Definition formulae for the proposed quantities are based on physical properties of diffuse sound field and psychoacoustic properties of human hearing. Variability of SIRs with respect to the measurement setup geometry shows that there is a possibility to apply the proposed factors in practice.

On the basis of the obtained results it can be concluded that most of the proposed quality factors meet the adopted assumptions. Small changes in SIRs result in a systematic decrease of values of individual parameters (except for F_{st} values of which increase). Only in three ($R_{\varphi EDT}$, $R_{\varphi st}$, $R_{\theta st}$) from the total number of ten parameters, the desired trend is not observed. However, to decide ultimately on usefulness of these inconsistent indices for the purpose defined in this paper, additional research and/or calculations would be necessary. Another problem is connected with interpretation of results which seem to be dependent on the type of room used in the experiment. Such inconsistency could be likely overcome by renormalisation of the proposed indicators with respect to a field dissipation measure and/or the reverberation time.

Therefore, several psychoacoustic tests should be performed in order to determine proper weights and coefficients for the proposed AQFs. Many different B-format SIRs need to be measured and compared before a wider application of the designed factors will be possible. Results of such experiments and measurements can be of help to designers of multi-channel sound systems and microphone recording techniques, as well as room acoustics consultants and architects.

Acknowledgments

This research was partly supported by AGH University grant no. 15.11.130.748.

References

- [1] W. Woszczyk, T. Beghin, M. de Francisco, D. Ko, *Proc. 127th AES Convention*, Audio Engineering Society, New York 2009, paper 7856.
- [2] T. Lokki, H. Vertanen, A. Kuusinen, J. Pätynen, S. Tervo, *Proc. of the International Symposium on Room Acoustics*, Australian Acoustical Society, Melbourne 2010.
- [3] P. Małeck, J. Wiciak, J. Wierzbicki, *Acta Phys. Pol. A* **123**, 1114 (2013).

- [4] J.H. Rindel, C.L. Christensen, *Proc. of the EAA Symposium on Auralization*, The European Acoustics Association, Espoo 2009.
- [5] P. Malecki, Ph.D. (in Polish), AGH University of Science and Technology, Krakow 2013.
- [6] S. Yokoyama, H. Yano, H. Tachibana, *Proc. INTER-NOISE*, Honolulu 2006.
- [7] S. Weinzierl, S. Feistel, C. Christensen, M. Vorländer, *Round Robin on Auralization*, EAA Joint Symposium on Auralization and Ambisonics, Berlin 2014.
- [8] P. Coverdale, S. Moller, A. Raake, A. Takahashi, *IEEE Signal Proc. Mag.* **28**, 91 (2013).
- [9] A. Raake, C. Schlegel, *Proc. of the ITG-Fachtagung Sprachkommunikation*, Aachen 2008.
- [10] ITU-R BS.1387-1: *Method for objective measurements of perceived audio quality*, 2001.
- [11] R. Nicol, L. Gros, C. Colomes, O. Warusfel, M. Noisternig, H. Bahu, *Proc. of the EAA Joint Symposium on Auralization and Ambisonics*, Berlin 2014.
- [12] S. Spors, H. Wierstorf, A. Raake, F. Melchior, M. Frank, F. Zotter, *Proc. IEEE* **101**, 1920 (2013).
- [13] F. Zotter, M. Frank, *J. Audio Eng. Soc.* **60**, 807 (2012).
- [14] A. Avni, H. Wierstorf, M. Geier, J. Ahrens, S. Spors, B. Rafaely, *J. Acoust. Soc. Am.* **133**, 2711 (2013).
- [15] J. Merimaa, V. Pulkki, *J. Audio Eng. Soc.* **53**, 1115 (2005).
- [16] N. Epain, C.T. Jin, F. Zotter, *Acta Acust. u. Ac.* **100**, 928 (2014).
- [17] T. Rossing, *Handbook of Acoustics*, Springer, Stanford 2007.
- [18] M. Long, *Architectural Acoustics*, Elsevier, San Diego 2006.
- [19] ISO 3382, *Acoustics – Measurement of the Reverberation Time of Rooms with Reference to Other Acoustical Parameters*, 1997.
- [20] L. Beranek, *Concert and Opera Halls*, Acoustical Society of America, New York 1996.
- [21] Y. Ando, *Acustica* **76**, 292 (1992).
- [22] T. Kamisiński, *Acta Phys. Pol. A* **118**, 78 (2010).
- [23] K. Kosała, Z.W. Engel, *Applied Acoustics* **74**, 1144 (2013).
- [24] A. Kulowski, *Applied Acoustics* **15**, 3 (1982).
- [25] M. Azevedo, J. Sacks, *Proc. of the EAA Joint Symposium on Auralization and Ambisonics*, Berlin 2014.
- [26] B.E. Goldstein, *Sensation and Perception*, Cengage Learning, Wadsworth 2010.
- [27] ANSI/ASA S1.42-2001, (R2011), *Design Response of Weighting Networks for Acoustical Measurement*, 2011.