Acoustics'08 Paris
June 29-July 4, 2008
www.acoustics08-paris.org

Assessment of simulation quality of three different auralization procedures

Stephan Paul

Lab. of Vibrations and Acoustics, Fed. Univ. of Santa Catarina, CTC, EMC, Campus Trindade, 88040 Florianópolis, Brazil
stephan.paul.acoustic@gmail.com

When acoustical problems occur in architectural spaces acousticians are often requested to help. The help given is often complicated by those involved having a very different knowledge in acoustics and by problems in communication between architects, musicians and acousticians. Auralization of sound sources is considered to be a powerful tool in solving the problem, making acoustic information accessible for listening. The research study here reported assessed the naturalness of the output of different simulation procedures by subjective assessment of similarity with a reference sound, considering different perceptual dimensions (localization, reverberation and timbre). The reference sound was obtained by binaural recording in a room that was auralized using ray-tracing, image sources or a hybrid pyramid tracing method. Two different types of sound sources, human speech and a guitar, were auralized by Marcelo Portela from Federal University of Santa Catarina. The results show that the subjective quality of the simulation procedure is dependent on the type of sound source and the perceptive dimension that was assessed. In general, the image source method performed best for localization of speech, whereas ray-tracing was the preferred method for the guitar sound. Overall image sources also outperform the other methods when good localization is required. Regardless of the source, the hybrid method was considered the best for the reverberation dimension.

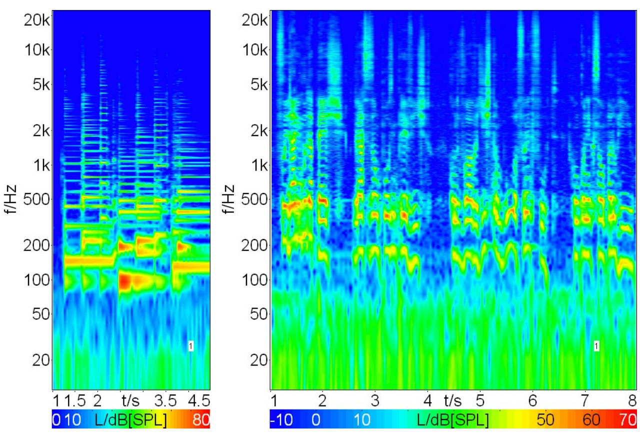
1 Introduction

Auralization is a powerful tool in processing acoustic effects, primary sound signals or means of sound reinforcement or sound transmission into an audible result. In this way a direct assessment of acoustic effects, sources, means of sound reinforcement etc. using our hearing sense is possible.

Nowadays, very different auralization procedures exist, Vorländer gives a good overview [6], and each has its strengths and weaknesses. The study presented in this article aimed to assess the quality of auralization results originating from an image source, ray-tracing and hybrid pyramid tracing, subjectively. This was carried out through direct comparison with binaural recordings, the latter made especially for this purpose in a real room using a manikin at the same position as the virtual listener in the auralization of the sources in the same room.

2 Auralization

Two different sources, a human male voice reading a sentence from Chico Buarque's "Budapest" and a short sequence played on a guitar, were recorded under anechoic conditions and auralized at two different positions in the geometric model of a real room.



(a) guitar (b) voice

Figure 1: Sonograms of the anechoic recordings of the guitar sequence and the human male voice

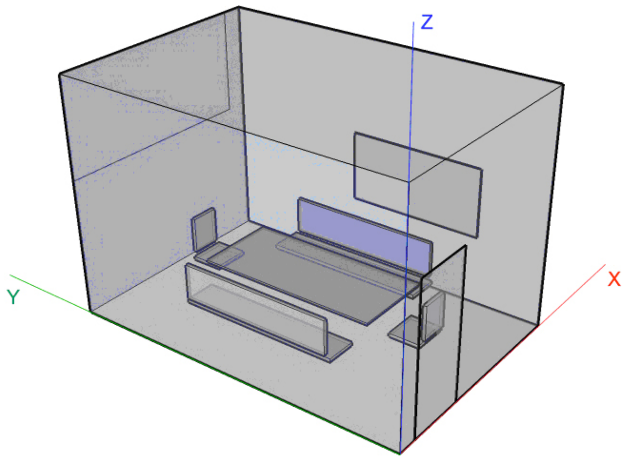


Figure 2: Geometric model of the room used in the auralization

Three different methods were used by Marcelo Portela at the Federal University of Santa Catarina to generate the room impulse response required for auralization of the different sources at different positions [5]. The first method was the image source method, the second classical ray-tracing and the third was a (hybrid) pyramid tracing method. The image source and ray-tracing methods used were implementations in a software program still under development [1]. The commercial software RAYNOISE was used for the pyramid tracing method. The impulse responses obtained in this way were combined with the head related transfer functions HRTFs which were kindly provided by Prof. Michael Vorländer from the Institute of Technical Acoustics at Aachen University. The auralization procedures used in this study, as well as the simplifications carried out and the assumptions made, are described in [4].

3 Subjective evaluation procedure

Using a scaling technique the results from the different auralization procedures for both sources, at two positions each, were compared to the binaural recordings of the same sources at the same positions. A continuous rating scale (line-scale or visual analogue scale) was used to assess the dimension very similar - very different, using verbal endpoints in Brazilian Portuguese as proposed by Leite&Paul [2]. For ease of use the length of the scale was fixed to ten cm.

Figure 3: The continuous rating scale with simple verbal endpoints

Subjects were required to place the letter corresponding to every auralized sound they heard on the scale according to their perception of similarity with the reference sound, this was the binaural recording. Figure 5 gives an example of a completed scale.

The procedure, except the rating scale itself, was implemented into a human computer interface, in order to allow for self-paced tests. The sounds were presented via Sennheiser HD 580 electrodynamic headphones.

In a pre-test ten persons evaluated the results of the three different auralization procedures of the two sources at two different positions. Subjects were asked to evaluate the general similarity or dissimilarity using the scale as given in Figure 3. It was found that this posed considerable difficulties to the subjects. It was therefore decided to refine the assessment procedure.

In new tests, the subjects were required to evaluate similarity/dissimilarity regarding three different concepts, these being (1) localization, this is the direction where the sound appears to come from; (2) reverberation, as related to the apparent volume of the room, and (3) the timbre.

Prior to the evaluation subjects had a short introductory lesson on auralization, although they still could not be considered to be familiar with the topic. All different concepts (localization, reverberation and timbre) were introduced to the subjects by audio examples given by the computer interface just before requiring the evaluation of the sound samples regarding the concept.

The assessment carried out in this way resulted in 2(sources)$\times$ 2(positions)$\times$3(concepts) = 12 completed scales as shown in Figure 5 with their respective marks. The auralized sound samples were presented using the letters A, B, and C as shown in Figure 4.

Thus, the subjects were aware of the position of the sound source. The letter A corresponded to the image source method (IS), B to the hybrid pyramid tracing method (PT) and C to common ray-tracing method (RT). Subjects marked their responses on the continuous rating scale given on a paper form where the letter R indicates the reference.

A total number of 35 subjects participated, mainly naive ones.

4 Results

Data were analyzed in two different ways, considering that interval-like and ordinal data were obtained. Regarding the (near)-interval data univariate outliers were identified and excluded using boxplots. In this way

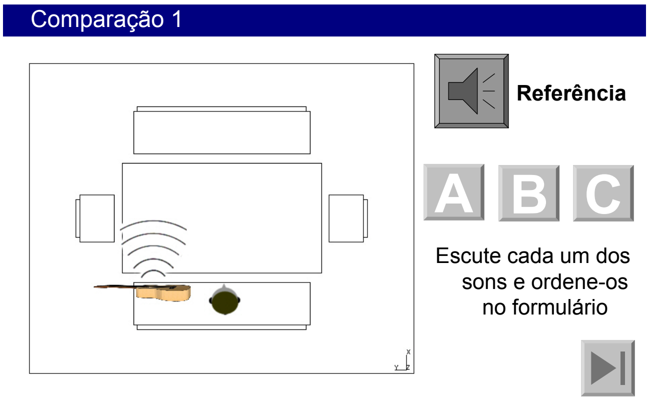


Figure 4: Example of a screen of the computer interface. The screen shows the auralized situation (guitar beside the listener) and the buttons for the reference sound and the three auralized sound samples A-C.

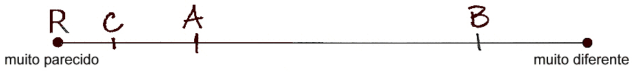


Figure 5: An example of a completed line scale

seven out of 420 completed scales were discarded, representing only 1.67% of the data.

The data were analyzed regarding their ordinal and their interval properties. This procedure was chosen because some subjects indicated that they had difficulties in properly assigning distances to their checkmarks.

4.1 Results considering ordinal data

From a first inspection of the rank data it was possible to conclude that the image source method performed best for localization of the human voice whereas the ray-tracing performed best regarding localization of the guitar. Even though the difference was small, this may indicate that the type of source plays a role. Without distinction of the source, the image source method was rated the best regarding quality of source localization.

Regarding reverberation as the dimension to be evaluated the participants ranked the pyramid-tracing as the best auralization algorithm. Regarding the timbre once again pyramid-tracing scored best, but the advantage over the other methods was extremely small.

Using a rank analysis and a Wilcoxon-Signed Rank Test the significance of the differences between the ranks given to the different auralization methods was assessed. The results are presented in Table 1.

The Wilcoxon-Signed Rank Test indicated that there are no differences between the ranks associated with image sources and ray-tracing for the guitar sound either for pyramid tracing or ray-tracing for the human voice and in general. The conclusions from the ordinal data are given in Table 2.

Table 1: Results of the Wilcoxon-Signed Rank Test. **y** indicates that there is a statistical significant difference ($p = 0.05$), **n** indicates that there is no such difference.

parameter	source	IS/PT	PT/RT	IS/RT
localization	guitar	y	y	n
	voice	y	n	y
	general	y	n	y
reverberation	guitar	y	y	n
	voice	y	y	n
	general	y	y	n
timbre	guitar	n	n	n
	voice	n	n	n
	general	n	n	n

Table 2: Results from ordinal data

		rank		
parameter	source	1	2	3
localization	guitar	IS & RT	PT	
	voice	IS	RT & PT	
	general	IS	RT & PT	
reverberation	guitar	PT	IS & RT	
	voice	PT	IS & RT	
	general	PT	IS & RT	
timbre	guitar	no classification		
	voice	no classification		
	general	no classification		

4.2 Results considering interval data

Additionally to the ranks that can be obtained from the scale interval, data are also obtained by transforming the position of the checkmarks into a numerical value. Beginning with 0 for very similar up to ten for very different, all checkmarks were transformed into numerical values with 0.5 precision. From histograms it was observed that the data for each scale are seldom close to a normal distribution. Therefore, the median was chosen as the localization parameter rather than the mean. Table 3 presents the median values obtained.

It can be observed that nearly all median values, with the exception of rating for the reverberation concept of the classical ray-tracing method, are lower than or equal to five, that is, they are considered more similar to the reference than different from the reference. However, this could be the result of an acquiescence bias.

Before establishing a ranking based on the median values a Friedman-test and a Kendall-W test were performed to check for differences considering $p = 0.05$. The results are given in Table 4.

From this data a significant difference was observed for the localization of the guitar sound regarding the image source and the classical ray-tracing method. The same is true for the reverberation dimension. Thus, the methods may be reclassified as shown in Table 5.

Table 3: Medians from interval data. The lower the median the more similar the auralization to the binaural recording.

parameter	source	IS	PT	RT
localization	guitar	2.5	5	2
	voice	2.5	4.5	4.5
	general	2.5	5	3.5
reverberation	guitar	3.25	2	5
	voice	5	4	6
	general	4.5	3	5.5
timbre	guitar	4	3	4.25
	voice	5	4.5	4.5
	general	4.25	3.5	4.5

Table 4: Results of the the Friedman-test and Kendall-W test. **y** indicates that there is a statistically significant difference ($p = 0.05$), **n** indicates that there is no such difference. The " " signs were used to indicate differences with respect to the analysis of the ordinal data.

parameter	source	IS/PT	PT/RT	IS/RT
localization	guitar	y	y	"y"
	voice	y	n	y
	general	y	n	y
reverberation	guitar	y	y	n
	voice	y	y	n
	general	y	y	"y"
timbre	guitar	n	n	n
	voice	n	n	n
	general	n	n	n

Table 5: Results from interval data

		rank		
parameter	source	1	2	3
localization	guitar	RT	IS	PT
	voice	IS	RT & PT	
	general	IS	RT & PT	
reverberation	guitar	PT	IS & RT	
	voice	PT	IS & RT	
	general	PT	IS	RT
timbre	guitar	no classification		
	voice	no classification		
	general	no classification		

For the timbre dimension once again no classification was possible.

On analyzing the comments of the subjects it must be acknowledged that they had two difficulties. The first was to assign distances from the endpoints of the scale for the placement of the check. Thus, the ordinal data may provide more stable results than the interval data. Also, the subjects reported that the independent evaluation of the three concepts was difficult, as they are not truly orthogonal, as previously indicated by Lokki [3]. However, they agreed that this was still easier than the general assessment required in the pre-tests. The timbre dimension posed particularly large difficulties to the subjects, probably due to the complexity of the concept of timbre.

5 Conclusions

Auralization is a powerful tool in making acoustics and acoustic interventions explainable to naive subjects. Different methods of auralization have been evaluated regarding their naturalness, that is, their similarity to binaural recordings made in the real room. The evaluation was based on paired comparisons with the binaural recordings.

All methods showed strengths and weaknesses according to the source and the perceptive dimension (localization, reverberation and timbre) evaluated. The results indicated that the subjective quality of the simulation procedure is dependent on the type of sound source and the perceptive dimension assessed. In general, the image source method performed best for localization of speech, whereas ray-tracing was the preferred method for the guitar sound. Overall, image sources also outperformed the other methods when good localization was required. Regardless of the source, the (hybrid) pyramid tracing method was considered the best for the reverberation dimension.

All of the subjective evaluations were found to be complex, especially when naive subjects are requested to provide interval data and to assess the timbre of the sounds presented.

Acknowledgements

I very gratefully acknowledge the work of Marcelo Portela from Federal University of Santa Catarina who carried out the auralization and helped with the subjective evaluation.

References

[1] M.H.A Gomes. Implementação de métodos para a simulação acústica e auralização de salas. *Acústica & Vibrações*, 38, 2007.

[2] Ricardo Penna Leite and Stephan Paul. Qualidade sonora: conceitos básicos ilustrados com um exemplo da área automotiva. *Acústica & Vibrações*, 37:28–37, 2006.

[3] Tapio Lokki and Hanna Järveläinen. Subjective evaluation of auralization of physics-based room acoustics modelling. In *Proceedings of the 2001 International Conference on Auditory Display*, Espoo, Finland, 2001.

[4] Marcelo Portela. Caracterização de fontes sonoras e aplicação na auralização de ambientes. Master's thesis, Federal University of Santa Catarina, 2008.

[5] Marcelo Portela and Stephan Paul. Auralização de voz e violão – uma abordagem comparativa de diferentes métodos de simulação. *Acústica & Vibrações*, 39, 2008.

[6] Michael Vorländer. *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality.* Springer, 1st edition, 2008.