An interactive auralization method using real-time sound sources

Josefin Lindebrink[a)]
Department of Acoustics, Tyréns AB, Peter Myndes Backe 16, SE- 118 86 Stockholm, Sweden

Jens Forssén[b)]
Division of Applied Acoustics, Department of Civil and Environmental Engineering, Chalmers University of Technology, SE-41296 Gothenburg, Sweden

During recent years, auralization methods have evolved towards using more interactive measures. The use of interactive elements, like navigation in static sound fields, has proven to be very significant in order to better integrate the listener with the simulated soundscape. In this study the possibility of engaging the user by actively contributing to the sound field is explored. Enabling the subject to act as a sound source and allowing communication within the environment, utilizing real-time synthesis of an acoustic environment's response. Auralization allows for a psychoacoustic evaluation of the acoustical space and therefore plays an important part in a wider understanding of different environmental characteristics. With an auralization framework adapting this kind of interaction, experience of the acoustical response is enabled and can thus be used as a tool in the process of subjectively assessing the acoustical space. Real-time convolution software implementing this mode of procedure has been designed. A subjective evaluation has been performed using a listening room equipped with an ambisonics multi-channel reproduction system, and a directional microphone with feedback control. Evaluation results indicate a positive response from the subjects to the added control over the simulated space.

[a)] email: josefin.lindebrink@tyrens.se
[b)] email: jens.forssen@chalmers.se

1 INTRODUCTION

Real-time auralization methods have developed vastly during recent years, implementing more interactive measures. Among these are applications allowing the subject to move in static sound fields or in real-time alter its acoustical characteristics, instantly hearing the effects[1-3]. Acoustic calculation software such as CATT-Acoustics is now also offering added features of auralization using interactive elements. It is no question that we are moving towards these kinds of solutions, activating and engaging the listener with the goal of reproducing realistic and immersive Virtual Acoustic Environments, VAE.

With auralization one can in an early designing stage assess subjective preference and suitability of acoustical environments to a specific purpose as well as health effects. Utilizing interactive measures, the question now arises as to what modes of interaction should be included in the auralization process, and how these methods could be beneficial, further improving the conditions for the subjective assessment. Several modalities of interaction should possibly be used, engaging the listener in more ways than one, thus resembling the control benefitted from real environments.

The alterations in the use of our own voice when subjected to different environments have been discussed in previous studies[4], suggesting that talking quality in enclosures could be assessed by experiencing the response of the room in form of reflected energy. As aspects of the speech production are affected, in ways such as adjustments in loudness due to background noise (i.e. the Lombard effect), or altering the speech rate, it implies that conscious or subconscious actions are made adapting to the surroundings. If these actions can be used in the room acoustical assessments of VAE, it offers a great way of integrating the listener with the environment, also allowing subjects to use their own voices as reference. It also makes it possible to perform virtual meetings, where several participants could communicate in the simulated environment, possibly making the simulation less obvious.

Methods of enabling musicians to play their instrument in VAE have previously been proposed,[5,6] as it has been found that musicians are well aware of the environment of which they play their instrument in, altering the way of playing depending on its response. Similarly, if aware of the changes in the way we use our voice in different environments, these alterations could possibly be used as a tool when assessing the room acoustical qualities of VAE, knowing our own voice and being able from a very early age to adapt the way of using this depending on the response of the surroundings.

Therefore, an application allowing the subject to act as a sound source has in this project been compiled and evaluated using a multi-channel reproduction system and a continuous microphone-feed from the listening room, controlling acoustical feedback of the loudspeaker-room- microphone system with necessary measures. Finally the application has been evaluated by a small listening test with the purpose of studying subjects’ response to the added interaction as well as possible effects on the perception of the simulated environment.

2 AURALIZATIONS IN THE ACOUSTICAL DESIGN PROCESS

Since auralizations can be used to make the transmission path audible of sound emitted in an environment, by a sound source to a receiver, it is a great tool to use in the acoustical design process, enabling possibilities of directly hearing and comparing environmental effects with and without measures of acoustic control.

The theoretical nature of traditional acoustics has made the possibilities and importance of acoustic design somewhat inaccessible to people outside of the acoustics-community, however using auralizations, direct access can be given to experiencing the possibilities of an adequate acoustical design. As Furlong et al. have discussed[3] simple solutions need to be applied, enabling an insight to the design process. It is also of importance that the simulated sound fields can be perceived as realistic and immersive. If this is achieved by adding interactive elements, it can allow the acoustical design to be considered at earlier designing stages, thereby influencing the final design of the space to a greater degree.

3 PSYCHOACOUSTICAL CRITERIA

When evaluating subjective perception of sound fields some physical parameters are of importance. These can be derived from the impulse response and are the sound pressure level of the direct sound, L_p, and the level, angle and arrival time of reflected sound components[7]. The impulse response contains information about early reflections as well as the reverberation tail, the later describing the sound energy distribution of the environment. From the RIR, several parameters of the simulated environment can be assessed.

There are many available parameters for assessing the subjective response to acoustical environments. The primary acoustical parameter has in studies proven to be the reverberation time[8,9], RT_{60}, mainly corresponding to the perception of room size. As the goal of this work has been to evaluate the effects of the added interaction, the perception of room size has been used as the main parameter of evaluation. Possibly giving an initial indication on what effects the added interaction has and if it is beneficial or detrimental to the conditions for assessing the simulated sound field. That is, a comparison have been made between only listening to auditory events occurring in the simulated environment and adding control by being able to make sounds and contribute to the sound field.

4 AURALIZATION FRAMEWORK

4.1 Methodology

To simulate the effects of speaking in a room, the source and receiver should be modelled at close distance when deriving the room impulse response (RIR), corresponding to that distance between the subject's mouth and ears. Having the sound source present at the time of auralization, the derived RIR used for simulation should not contain the direct pulse, as this is created by the subject in the listening room. The RIRs should therefore be edited, removing the direct pulse before auralization, only producing the reflected energy of the simulated environment through the reproduction system.

As communicational purposes are of interest, a system allowing several simultaneous participants is preferable, enabling these to communicate within the virtual environment. Therefor a multi-channel loudspeaker system has been adapted and ambisonics used for the decoding of the distributed sound field. To handle acoustic feedback between the loudspeaker system and the microphone, a narrow pick-up pattern (i.e. similar to the shotgun directivity) of the microphone has been necessary to use, and acoustic feedback control has been applied using automatic parametric-filter based equalization.

With the purpose of only studying one interactive parameter, only static source and receiver positions have been used, whereby we are able to calculate the impulse responses offline. It also means that the application can be implemented using an ordinary personal computer, without the need for any external digital processors.

4.2 Demands on an Interactive Auralization Method Utilizing Real-time Sound Sources

For a real-time audio application, it is necessary to keep latency by signal processing low. The chain of processing between the time the subject makes a sound, to the time the convolved response of the rooms reaches the subject's ears, should correlate to that between the direct sound and the early reflections of the impulse response, keeping additional time below the audible limit of delay, preferably below 20-30ms[4]. Perceived time delay or echoes would deteriorate the perception of an immersive, realistic sound field. Coloration by the listening room or equipment used for reproduction and recording should also be avoided, keeping the sound quality intact.

The convolution processing between the sound source and the calculated impulse response will be one of the most demanding signal-processing operations. Choosing the right method for performing this convolution is important; weighting time delays with possible lengths of RIRs used ensuring no audible cuts of the reverberation decay. One should be able to use RIRs at least a few seconds long.

Furthermore, to ensure a correct reproduction level of the reflections, the sound pressure level relationship between the direct sound and the reflections reproduced by the loudspeaker system needs to be known. Calibration of the loudspeaker amplification should be performed to ensure right reproduction level at the receiver point in the listening room. Since the distance between the subject's mouth and ears is very small, there will be a large dynamic signal difference between the direct sound and the later parts of the impulse response, which in turn would require a sufficiently large dynamical range of the digital output file. Approximations concerning the RIR, discussed below, have been made to handle this issue, avoiding resolution problems with associated noise-like artefacts from occurring during the auralization. Furthermore, a feedback control system may be needed to avoid acoustical feedback of the loudspeaker-room-microphone system, which can influence the simulated room reverberation decay.

The listening room used for auralization needs to be relatively damped, not adding any coloration or affecting the simulated environment. The listening room should have a shorter reverberation time than the modelled environments and no pronounced early reflections. Also the background noise level has to be low in comparison to the sound pressure level of the late reverberation tail of the modelled enclosure[3].

4.3 Derivation and Preparation of Impulse Responses

Since using static source and receiver positions, the calculation of the impulse responses can at this point be performed offline. In this case the software CATT-Acoustics and its module for derivation of impulse responses, TUCT have been used. The TUCT-module utilizes cone tracing for calculation of the reverberation tail, and image source modelling for the initial early reflections. The rooms have been modelled in 3D graphical software, defining room geometry and materials. As the 3D model is exported into CATT-Acoustics, absorption, scattering and transmission coefficients are set. The source and receiver positions have for the subjective testing been defined using a 0.5 m separation distance to improve on the dynamical range issue, while at the same time giving a reasonable approximation of the reflected paths to the subject. The source has been modelled using the talker source directivity data included in CATT-Acoustics directivity databank. The impulse responses have been retrieved in the B-format due to the chosen reproduction system. As there is a requirement on a maximum latency of the signal processing, the initial time delay gap between the direct sound and the early reflections can be used as a buffer to compensate for any latency when also removing this from the RIR.

4.4 Application Architecture

The auralization application was implemented using the graphical programming software Pure Data, Pd. The application contains a loading function of the 4-channel RIR imported as combined wav-files. The RIR together with a continuous signal from the microphone in the listening room are sent to the convolution algorithm, in this case Ben Saylor’s function *partconv~*, of the Pd extended library. Utilizing blocked FFT-convolution, the convolution process together with a time efficient execution algorithm is able to perform convolution continuously without excess delays as long as the block-size of the convolution is sufficiently small. The chosen algorithm can divide the signal into a maximum of 256 blocks; enabling a RIR up to 3 s long to be used with time latency kept below that of audible delay (corresponding to the signal processing time of 1 block). For a shorter RIR, the block-size can be further reduced as well as the inherent time delay.

An ambisonics decoder is utilized, also obtained from the Pd extended library compiled by Thomas Musil and the Institut für Elektronische Musik und Akustik in Graz, Austria. The decoder is used for the distribution of the convolved signal to the loudspeakers, able to handle higher orders of ambisonics, producing both two and three-dimensional sound fields. Number of loudspeakers and their respective relative angle is set within the application. Time delay units as well as separate gain units are applied to each loudspeaker signal, giving opportunity to calibrate the loudspeaker system in-software if necessary. The application architecture is illustrated in Figure 1.

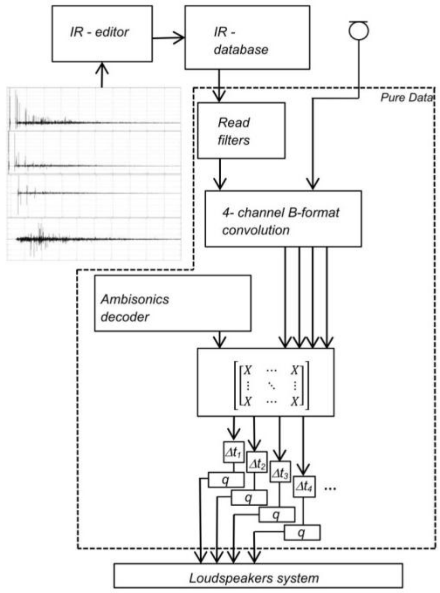

Figure 1 – Application architecture.

4.5 The Sound Design Lab, listening environment and reproduction system

The application has been implemented in the Sound Design Lab, see Figure 4, at the University College of Arts, Crafts and Design in Stockholm, which is a lab used in collaboration with the Acoustics Department of Tyréns AB. The lab is equipped with a 6 to 8-channel loudspeaker system utilizing 18 mid- and high range loudspeakers combined with 4 subwoofers, enabling a hemispheric sound reproduction. It has a low reverberation time of about 0.2 s in the overall frequency spectrum, and measures for controlling resonance effects of low frequency modal patterns have been applied. The background noise level is kept low. A narrow-pick up microphone has been suspended from the ceiling at a 45° angle in the median plane relative to the audience seating. Acoustic feedback control has been implemented using an automatic parametric-filter based equalizer.

5 SUBJECTIVE TESTING

The application was evaluated in the Sound Design Lab by a group of 9 participants, 3 female and 6 male, all without previous knowledge of acoustics. The participants performed the test individually, with only a technician present to handle the application. The room was kept dark during testing to avoid visual influence.

The test was arranged in two sections, the first dealing only with the real-time auralization, enabling the subject as a sound source. The subjects were asked to judge perceived size of the simulated environment with the help of depicted venues of different volumes and with different characteristics. In section two a comparison was made between using a pre-convolved female talker sound source and the real-time sound source of the subject, these presented separately. Again the perceived size was used as a parameter of judgement. Additionally, the subjects were asked to judge the tonal character of the room, if sounding “harsh” or “soft” and to what extent they perceived the environment being realistic or authentic. When subjects were allowed to make sounds in the environment, they were free to make any sound they felt necessary for assessment, i.e. no manuscript was provided. Many adapted to clapping hands, speaking simple sentences and making short vocal transient sounds. The subjects also communicated with the technician during testing. The tests were concluded in a shorter interview where the subject could express their thoughts on the experience of the simulated environments and the effects of the interaction.

In the tests, five environments where used, whereof four were modelled in CATT-Acoustics and one was measured, the great hall of People´s Palace in London[10]. The reverberation times ranged from 0.4 to 2.1 s, with models shown in Figure 5(a)-(e).The rooms were modelled using a 0.5 m distance between the source and receiver. The listener was placed at 0.5 m distance from the microphone in the listening room, as depicted in Figure 2.

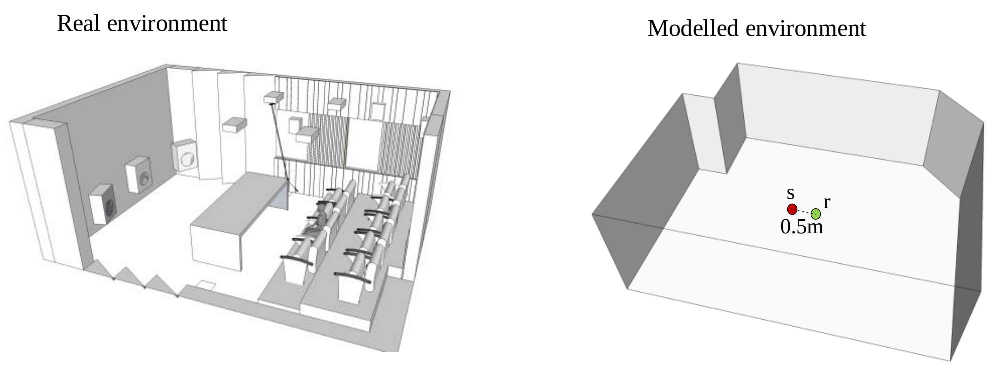

Figure 2 – Source and receiver positions, both in real, i.e. the listening room and in the simulated environment.

When evaluating the results, subjects’ correlation to large, medium sized and small rooms was compared with the relative reverberation time of the simulated environments as can be seen in Table 1. From these results, it seems that the relative perceived size correlates quite well with the RT_{60} of the simulated environments, having the majority of the subjects choose rooms with corresponding characteristics as the simulated enclosure.

Table 1 – Subject's response in section one, perceived size of the environment compared with reverberation time, RT60.

	RT_{60}, [s]	**Approximate room volume, [m^3]**	**Mean absorption, octave bands 500 – 2000Hz**	**Perceived size**		
				Large	**Medium**	**Small**
Great Hall	2.1	≥15000	-	9	0	0
Opera Hall	1.4	15000	35%	7	0	2
Open office space	0.4	450	43%	0	4	5
Lecture room	0.4	400	35%	0	5	4

In section two of the tests, comparing auralization using the pre-convolved and the real-time sound source, again the results seems to correlate sufficiently with the reverberation time. This time the subjects had to judge the rooms as either large or small. Only smaller variations of the subjects' response can be seen between the different auralization methods, shown in Table 2.

Table 2- Subject's response in section two, perceived size of the environment using a pre-convolved sound source and a real-time sound source, compared with reverberation time, RT60.

	RT_{60}, [s]	**Approximate room volume, [m^3]**	**Mean absorption, octave bands 500 – 2000Hz**	**Perceived size**			
				Pre-convolved sound source		**Real-time sound source**	
				Large	**Small**	**Large**	**Small**
Opera Hall	1.4	15000	35%	8	1	8	1
Canteen/cafeteria	0.4	3600	45%	4	5	3	6
Open office space	0.4	450	43%	1	8	1	8
Lecture room	0.4	400	35%	1	8	2	7

As results show in Figure 3, the subjects tended to judge the tonal character of the environment as softer when utilizing a real-time sound source.

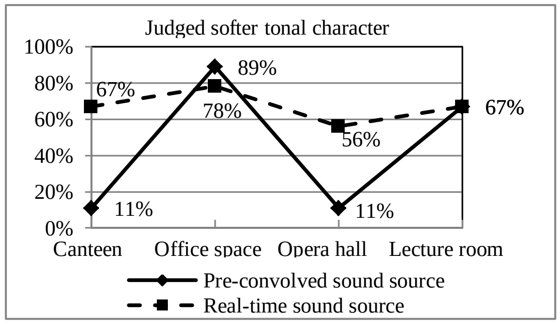

Figure 3 – Amount of subjects judging the tonal character as soft, results shown for each simulated environment.

During a shorter interview after completed evaluation, an overall preference was expressed of using the real-time application when performing these tasks, combined with a higher rating of authenticity than when utilizing a pre-convolved sound source.

6 CONCLUSIONS & FUTURE WORK

An auralization method utilizing real-time sound sources has in this project been implemented and evaluated. The application, compiled in the open source software Pd, is at this point able to auralize rooms with a room impulse response (RIR) length up to 3 s, without noticeable delay. At present, only static source and receiver positions are used, calculating and preparing the RIR offline. The application has been implemented in a controlled listening environment using an ordinary personal computer, a narrow-pick up microphone and a multi-channel loudspeaker system. Small latencies due to signal processing could be perceived as lagging, directly deteriorating the realistic impression of the simulation. A sufficiently small latency was achieved by using a default block-size corresponding to latency well below the audible delay limit.

Results from subjective evaluation tests indicate that sensing basic acoustical characteristics like room size seems not substantially affected by the added interaction. The difference in judged tonal character of the rooms could possibly be due to a psychological response to the added control. It can however not be excluded that the sound source used for the pre-convolved auralization could be a contributing factor.

A majority of the subjects responded that they appreciated experiencing the simulated sound fields by real-time auralization of their own voice. Combined with high ratings of authenticity, the application seems plausible to use in future projects. Although the subjects performed the tests individually, enabling several participants is called for, enabling a more natural communication situation in the simulated environment. When demonstrating the application to a group of participants, having the application running in the background during conversation, an adaption to the space could be sensed, creating awareness of the acoustical response, apparent when there was a transition to another simulated room.

The tests conducted in this project should only be seen as an initial indicator of the effects of utilizing real-time sound sources. Additional studies are however needed to draw any further conclusions as for the impact of the added interaction to the conditions for assessing the sound field.

The auralization application is today being installed at the University College of Arts, Crafts and Design and will be used by the consultants at Tyréns AB as well as by students and possibly research groups at the university.

Plans for the future include adding pre-convolved sound sources to be used simultaneously with the real-time ones, i.e. reproducing more complete sound fields including environmental sounds, background noises and other types of sources. Integrating this method with other forms of interaction should also be studied, in the hopes of closing in on the goal of reproducing realistic and immersive sound fields.

7 ACKNOWLEDGMENTS

This project was done as collaboration between Chalmers University of Technology and the Department of acoustics at Tyréns AB. The authors would like to thank these institutions as well as the University College of Arts, Crafts and Design.

8 REFERENCES

1. Savioja, J. Huopaniemi och T. Lokki, ”Creating Interactive Virtual Environments,” *J. Audio Eng. Soc*, Vol. 47, No. 9, 1999.

2. T. Lentz, D. Schröder, M. Vorländer och I. Assenmacher, ”Virtual Reality System with Integrated Sound Field Simulation and Reproduction,” *EURASIP Journal on Advances in Signal Processing*, 2007.

3. D. Furlong, M. Doyle, E. Kelly, C. MacCabe och R. MacLaverty, ”Interactive Virtual Acoustics Synthesis System for Architectural Acoustics Design,” i *AES Convention 93*, San Francisco, 1992.

4. R. Appel och G. Beerends, ”On the Quality of Hearing One´s Own Voice,” *J. Audio Eng. Soc.*, vol. 50, nr 4, 2002.

5. A. Gade, ”Investigations of Musicians´ Room Acoustic Conditions in Concert Halls, Part I.,” *Acustica*, vol. 69, 1989.

6. K. Ueno och H. Tachibana, ”Experimental study on the evaluation of stage acoustics by musicians using a 6-channel sound simulation system,” *Acoust. Sci. & Tech.*, vol. 24, 2003.

7. M. Kleiner, Audio Technology & Acoustics, 2nd ed. red., Gothenburg: Division of Applied Acoustics, School of Civil and Environmental Engineering, Chalmers University of Technology, 2008.

8. B. Shinn-Cunningham, ”Acoustics and perception of sound in everyday environments,” i *Proceedings of the 3rd Int. Workshop on Spatial Media*, Aisu-Wakamatsu, Japan, 2003.

9. S. Hameed, J. Pakarinen, K. Valde och V. Pulkki, ”Psychoacoustic Cues in Room Size Perception,” i *AES 116th Convention*, Berlin, Germany, 2004.

10. Stewart, Rebecca and Sandler, Mark. *"Database of Omnidirectional and B-Format Impulse Responses", in Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2010),* Dallas, Texas, March 2010.

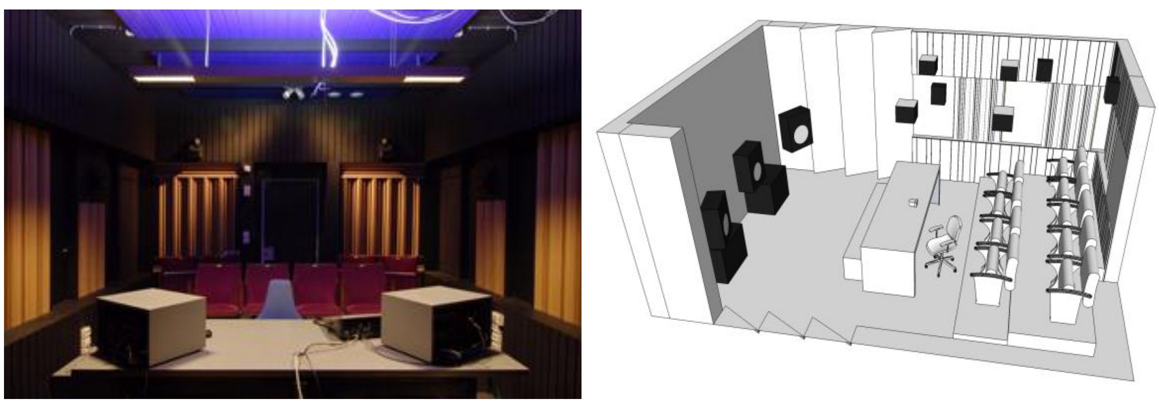

Figure 4 - the Sound Design Lab.

Figure 5a) - Opera hall

Figure 5b) - Lecture room

Figure 5c) - The Great Hall of People´s Palace[10]

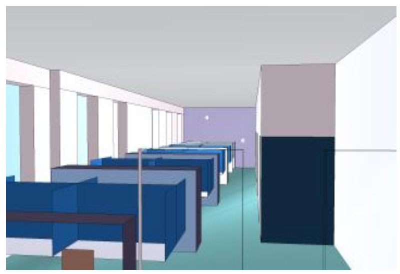

Figure 5d) - Small open office space

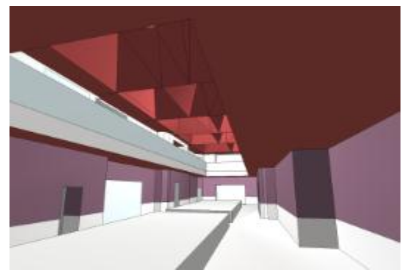

Figure 5e) – Canteen/Cafeteria

Figure 5 – Modeled and measured environments used for testing.