Real-time auralization of room acoustics for the study of live music performance

Sebastià V. Amengual[1], Dustin Eddy[1], Malte Kob[1], Tapio Lokki[2]
[1] *Hochschule für Musik Detmold, 32756 Detmold, Deutschland, Email: {amengual, kob}@hfm-detmold.de*
[2] *Aalto University School of Science, 02150 Espoo, Finland, Email: tapio.lokki@aalto.fi*

Introduction

Studying the influence of room acoustics on musicians requires the possibility of recording a live performance in different acoustic conditions in order to analyze the performance using either perceptual methods or extraction of music features using signal analysis. Due to the great amount of variables present in the acoustics of a room, it is necessary to have a comprehensive set of rooms with different conditions available to perform experiments, which is rarely the case. In addition, the musicians need to perform in every different room, increasing the logistics and procedure complexity. This problems can be solved by using a virtual acoustic environment which provides modifiable room acoustic conditions without the necessity of moving into different rooms, allowing a rapid change of acoustics for direct comparison and detailed control of the acoustic conditions.

This paper presents the implementation of a real-time auralization system of static acoustic scenes in which a musician can perform in real-time. The system aims at a faithful reproduction of room acoustics providing a framework for the study of live music performance in different acoustic conditions.

Methodology

The main idea of this implementation is to faithfully reproduce the reflections of a room while a musician performs inside a surrounding loudspeaker set-up. Spatial room impulse responses are measured in different rooms using a microphone array and a sound source on stage. Using the Spatial Decomposition Method (SDM) [1], associated impulse responses for an arbitrary loudspeaker set-up are generated. The sound of a musician performing in a quasi-anechoic room equipped with a surrounding spherical loudspeaker set-up is captured with close miking and convolved with the decomposed impulse responses. The convolved sound is then played back, reproducing the sound reflections in real time.

The following sections give a closer look at every step in the auralization process. A block diagram including all the operations is depicted in Fig. 1.

Measurements

A loudspeaker and a microphone array are arranged on stage in a similar position as a player would perform. The loudspeaker is at the instrument position and the microphone array is in place of the player's head. In order to improve the realism of the auralization, it is convenient to use a sound source with radiation characteristics similar

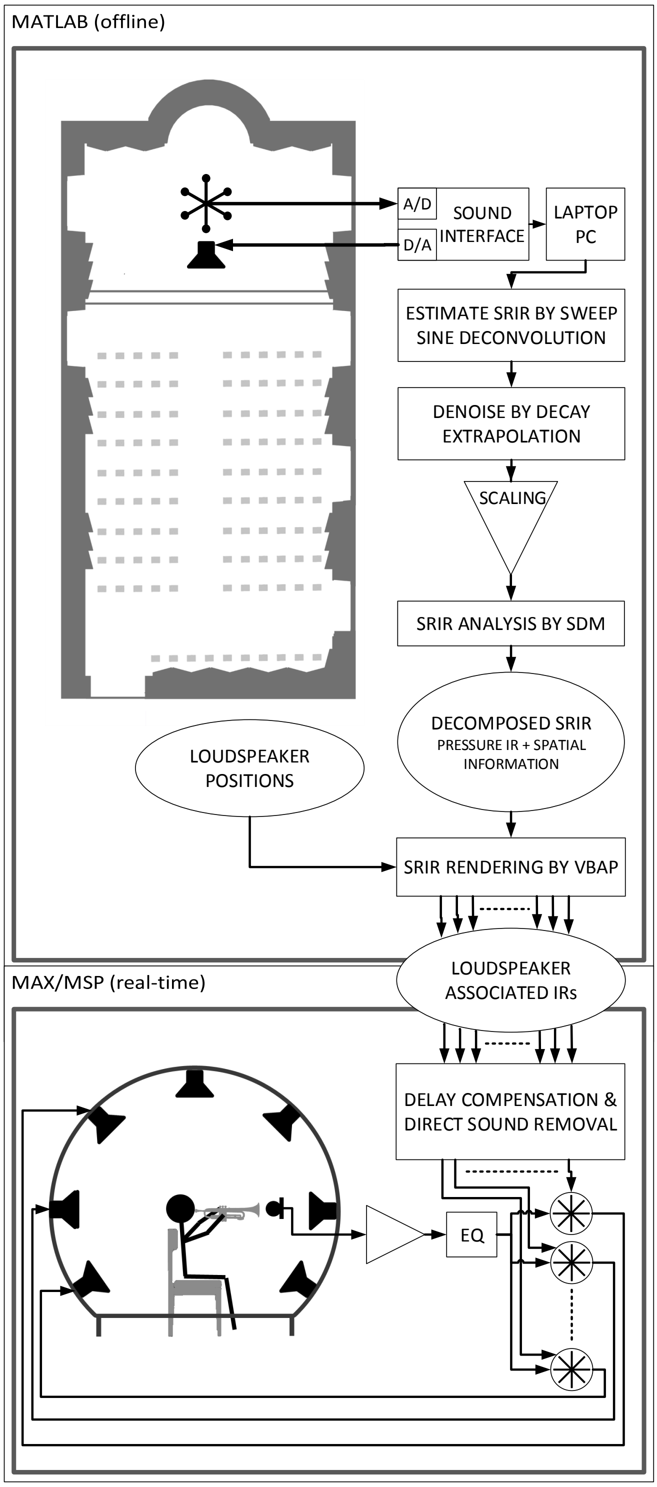

Figure 1: Main operations in the auralization process.

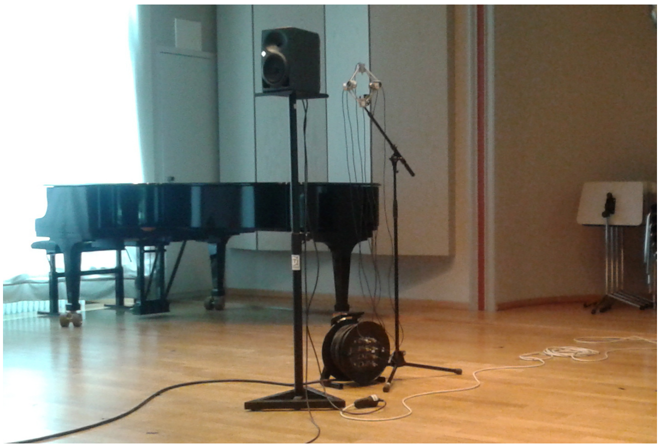

Figure 2: SRIR measurement set-up on stage

to the target instrument [2]. In this article the procedure is described for the auralization of a trumpet. It is important to note that the selection of the sound source, as well as the positioning on stage, would be different for every instrument.

The source used is a Neumann KH120A studio monitor. The microphone array is composed by 6 omnidirectional matched microphones (NTi M2010) placed orthogonally in a three-dimensional space at 10 cm apart.

A spatial room impulse response (SRIR) composed by 6 individual room impulse responses (RIR) is obtained by deconvolution of a reproduced sine sweep [3]. The SRIR are synthetically extended by extrapolating the energy decay in several frequency bands, as described by Cabrera *et al.* in [4] and normalized to ensure that the source level is the same in the different measured rooms. The measurement operations are performed using the ITA Toolbox for Matlab [5].

Spatial analysis

Using the SDM for impulse responses (IR) the SRIR is analyzed and decomposed as a succession of acoustic events with an associated direction of incidence for every sample. The working principle of SDM relies on the analysis of time differences between the arrival of sound events while assuming that an IR can be decomposed as a set of plane waves. A short moving window is applied to the IR estimating the direction of the arrival of sound at every time step (see Fig. 3). As a result, a SRIR is expressed as a pressure IR with associated spatial information. This implementation follows the one proposed by Tervo et al. in [1].

Spatial Synthesis

Spatial synthesis consists of creating a set of loudspeaker impulse responses (LIR) that can be later used for convolution with a live signal (or a recording). Since a decomposed SRIR is expressed as a set of images sources, implementation using Vector Base Amplitude Panning (VBAP) [6] is straightforward, and an individual source is rendered at the corresponding location of every time step. The advantage of VBAP is the possibility of using an arbitrary loudspeaker set-up that can be changed depending on the nature of the application. However, other methods such as Ambisonics or Near Loudspeaker Synthesis [7] can be considered as well in certain loudspeaker layouts.

Real-time Convolution

For real-time convolution of the rendered LIR and the sound produced by a musician, we developed a Max/MSP patch which uses the HISSTools external multiconvolve∼ object as a zero latency convolution engine [8].

In the auralization, the direct sound is generated by a musician, thus, it is necessary to remove the direct sound from the LIR before the convolution. In addition, the delay introduced by the convolution engine e.g. AD/DA conversion, sample buffers, etc., needs to be compensated accordingly to keep the same temporal relation between the incidence of the direct sound and the rendered room reflection on the musician's ears. In order to render the floor reflection, it is necessary to achieve a delay smaller or equal to approximately 8 ms. These operations are performed by the program when a new set of LIRs is loaded.

The modified LIRs are convolved in real time with the sound produced by the musician. For that we used a Schoeps CCV4 directional microphone very close to the instrument.

Live signal conditioning

The live input signal needs to be conditioned in order to reproduce the room reflections with appropriate spectral and level characteristics. Here, the microphone is in the near-field of the instrument (in front of the trumpet bell), and therefore the frequency characteristics of the microphone are affected. To correct this, two different musicians are recorded playing a chromatic scale in an anechoic chamber using the mentioned microphone in addition to an extra microphone at a distance of 2 meters. The spectral differences are obtained by comparing the spectra of the microphones. Using a parametric equalizer the spectral differences are corrected [9].

The process is as follows for the level calibration. First, a musician plays a fragment of music inside the listening room. This fragment is recorded near the player's ear at the listening position by an NTi M2010 measurement microphone. Then the recording is convolved with the part of the impulse response that is not used in the auralization - the direct sound of a measurement - and played back through the system. To find the input gain of the live signal, the level of the real recording and the level of the playback are matched.

Reproduction set-up

The listening room is a mixing studio of the University of Music Detmold with existing acoustic treatment complemented with removable absorption panels. The dimen-

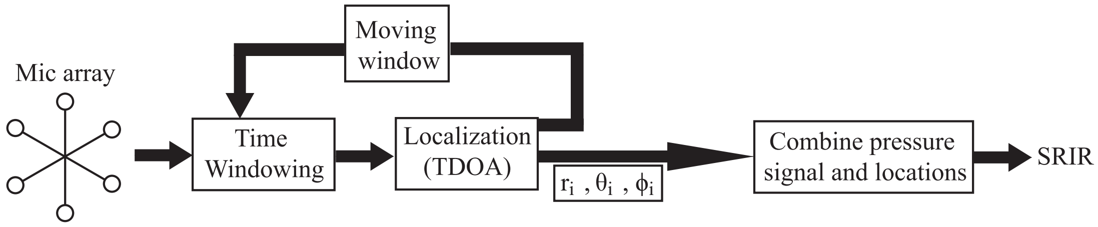

Figure 3: Overview of the operations for the analysis of a room impulse response using the Spatial Decomposition Method

Figure 4: Musician performing with virtual acoustic conditions

sions of the room are 5.6x4x2.8 m. Although it presents some low frequency modes, especially in the vertical dimension, the response of the room is predominantly anechoic with reverberation times equal to or lower than 0.1 seconds at mid and high frequencies.

The computer used for auralization is a Toshiba Satellite P50-A-12Z equipped with a quad-core CPU (Intel Core i7-4720HQ) and 8GB of RAM memory running Windows 8.1. The sound interface is an RME Madiface XT connected to an SSL Alpha Link AD/DA converter using a MADI connection. The I/O buffers are set to 64 samples. The overall delay of the system (input to listening position) is 7.6 ms, and is quickly measured before every use of the system using an acoustic analyzer with time delay measuring function (NTi XL2).

The loudspeaker set-up is composed of 13 Neumann KH120A studio monitors arranged in 3 rings of 4 loudspeakers at different elevations and a top loudspeaker, all at the same distance to the listening position (see Fig. 4. The loudspeakers are mounted on a spherical frame, and the set-up can be modified easily. The positions of the loudspeakers are collected in Table 2. The frequency response of the loudspeakers is flat within 2 dB between 50 Hz and 20 kHz.

Evaluation

An approach to carry out an objective evaluation of the auralization is to compare impulse responses of a real room and its auralization. The impulse responses from the auralization are obtained by feeding the input of the system with a swept sine and recording the system output at the listening position with an omnidirectional microphone. Note that the measurements contain only reflections and not direct sound, as it is intended to be created by a live musician.

The results of the comparison are included in Fig. 5 and they suggest that the time-frequency characteristics of the auralization match very well to the original room in a range up to 4000 Hz. Above this frequency range there are important deviations in the time-frequency properties of the auralization.

The lack of energy at high frequencies could be due to two reasons: a non-omnidirectional response of the measurement microphone and a non-coherent addition of the loudspeaker triplet signals that generate every single wavefront [10, 11]. In addition, due to the small wavelength at high frequencies, a small misalignment between the microphone position and the center of the listening environment could lead to large deviations in the results. Alternatives to correct these amplitude deviations at high frequencies are presented in [7] and [10] respectively, and they will be considered in the future to improve the performance of the present implementation.

The increase of reverberation time at high frequencies is a known feature of SDM, caused by the decomposition of the pressure impulse response as a set of image sources. Very fast transitions in the impulse responses lead to an increase of high frequencies, which is more relevant at the later part of the impulse response [1].

Conclusion

A system which allows real-time auralization and interaction of musicians with static acoustic scenes is presented. Although the auralized scenes present deviations from the target scene at high frequencies the system proved appropriate for the auralization of small instruments with limited frequency range. At the time of writing this paper a first trial experiment using the system for the study of trumpet live performance has been successfully completed.

Further upgrades of the auralization process include the correction of the non-omnidirectional response of the measurement microphones and a comparison with different spatial audio reproduction techniques in order to achieve the best results with the present hardware set-up.

Frequency	63	125	250	500	1000	2000	4000	8000	16000
T30	0.41	0.23	0.10	0.09	0.08	0.06	0.05	0.04	0.04

Table 1: Estimated reverberation time (T30) of the listening room averaged in octave bands.

Channel	1	2	3	4	5	6	7	8	9	10	11	12	13
Azimuth	225°	225°	225°	135°	135°	135	0°	45°	45°	45°	315°	315°	315°
Elevation	0°	-35°	45°	-35°	0°	45°	90°	0°	-35°	45°	0°	-35°	45°

Table 2: Positions of the loudspeakers in the reproduction set-up

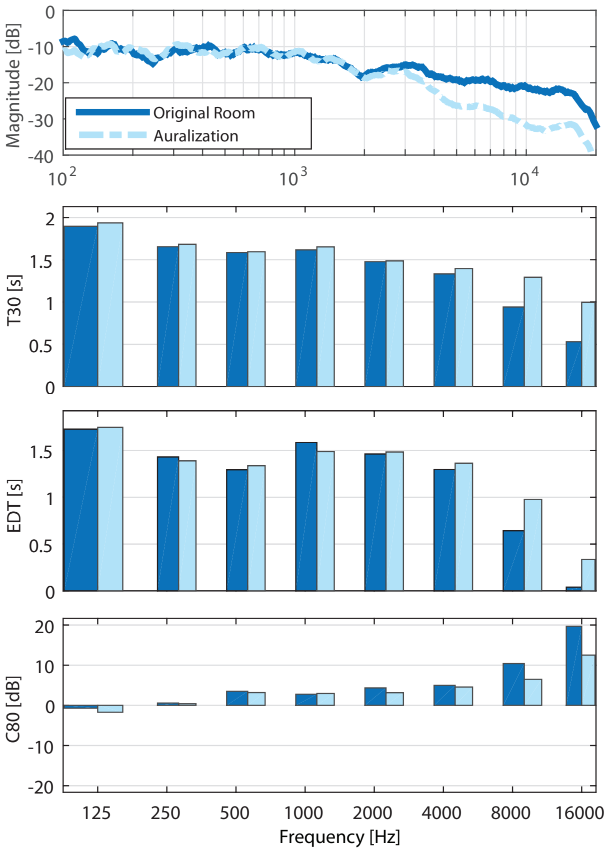

Figure 5: Impulse response magnitude spectrum and monaural parameters of a real room and its auralization.

Acknowledgements

The authors wish to thank colleagues Timo Grothe and Martin Schneider for providing directivity data of the trumpet and the Neumann studio monitor. The work presented in this article has been funded by the European Commission within the ITN Marie Curie Action project BATWOMAN under the 7th Framework Programme (EC grant agreement no. 605867).

References

[1] Tervo, S.; Pätynen, J.; Kuusinen, A.; Lokki, T.: "Spatial Decomposition Method for Room Impulse Responses," *J. Audio Eng. Soc.*, vol. 61, pp. 17-28, 2013.

[2] Otondo, F.; Rindel, J.: "The Influence of the Directivity of Musical Instruments in a Room", *Acta Acustica United with Acustica*, Vol. 90, no. 6, pp. 1178-1184, 2004.

[3] Farina, A.: "Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique", *108th AES Convention*, Paris, France, 2000.

[4] Cabrera, D.; Lee, D.; Yadav, M.; Martens W. L.: "Decay Envelope Manipulation of Room Impulse Responses: Techniques for Auralization and Sonification", *Proceedings of ACOUSTICS 2011*, Gold Coast, Australia, 2011.

[5] Dietrich, P.; Guski, M.; Pollow, M.; Sanches Masiero, B.; Müller-Trapet, M.; Scharrer, R.; Vorländer, M.: "ITA Toolbox - An Open Source MATLAB Toolbox for Acousticians", *DAGA 2012: 38. Deutsche Jahrestagung für Akustik*, Darmstadt, Germany, 2012.

[6] Pulkki, V.; "Virtual Sound Source Positioning using Vector Base Amplitude Panning," *J. Audio Eng. Soc.*, Vol.45, pp. 456-466, 1997.

[7] Tervo, S.; Pätynen, J.; Kaplanis, N.; Bech, S.; Lydolf, M.; Lokki, T.: "Spatial analysis and synthesis of car audio system and car cabin acoustics with a compact microphone array," *J. Audio Eng. Soc.*, vol. 63, no. 11, pp. 914-925, 2015.

[8] Harker, A.; Tremblay, P. A.: "The HISSTools Impulse Response Toolbox: Convolution for the Masses," *In ICMC 2012 : Non-cochlear Sound. The International Computer Association*, pp. 148-155, 2012.

[9] Rämö, J.; Välimäki, V.: "High-Precision Parallel Graphic Equalizer",*IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 22, no. 12, 2014.

[10] Laitinen, M.; Vilkamo, J.; Jussila, K.; Politis, A.; Pulkki, V.: "Gain normalization in amplitude panning as a function of frequency and room reverberance," *AES 55th International Conference: Spatial Audio*, 2014.

[11] Pätynen, J.; Tervo, S.; and Lokki, T.: "Amplitude panning decreases spectral brightness with concert hall auralizations," *AES 55th International Conference on Spatial Sound*, Helsinki, Finland, August 27-29, 2014.